

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 68.5 seconds
(Without alignments)
9697.360 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MQELSTCPFLCLLRPFCSA.....WVHQIALMEVLGCEAQQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04:*

- 1: geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB ID	Description
1	12410	100.0	2351	Aap50059 Human fac
2	12410	100.0	2351	Aaw11329 Native Fa
3	12410	100.0	2351	Aaw46245 Human fac
4	12410	100.0	2351	Aaw44373 Human fac
5	12410	100.0	2351	Aay52537 Human fac
6	12410	100.0	2351	Ab59528 Amino aci
7	12406	99.9	2351	Aaw11416 Active Fa
8	12406	99.9	2351	Aaw11343 Active Fa
9	12406	99.9	2351	Aaw11435 Active Fa
10	12405	99.9	2351	Aaw11445 Active Fa
11	12405	99.9	2351	Aaw11398 Active Fa
12	12405	99.9	2351	Aaw11425 Active Fa
13	12405	99.9	2351	Aaw11362 Active Fa
14	12405	99.9	2351	Aaw11461 Active Fa
15	12405	99.9	2351	Aaw11387 Active Fa
16	12405	99.9	2351	Aaw11419 Active Fa
17	12404	99.9	2351	Aap81113 Factor VI
18	12404	99.9	2351	Aap80659 Sequence
19	12404	99.9	2351	Aaw11408 Active Fa
20	12404	99.9	2351	Aaw11427 Active Fa
21	12404	99.9	2351	Aaw11347 Active Fa
22	12404	99.9	2351	Aaw11332 Active Fa
23	12404	99.9	2351	Aaw11437 Active Fa
24	12403	99.9	2351	Aaw11371 Active Fa
25	12403	99.9	2351	Aaw11396 Active Fa

ALIGNMENTS

RESULT 1
AAP50059
ID AAP50059 standard; protein; 2351 AA.

XX AAP50059;

DT 27-OCT-1991 (first entry)

DE Human factor VIII.

KW Human factor VIII.

OS Homo sapiens.

PN EP160457-A.

PD 06-NOV-1985.

PF 18-APR-1985; 85EP-00302734.

PZ 20-APR-1984; 84US-00602312.

XX (GETH) GENENTECH INC.

PA Capon DJ, Vehar GA, Lawn RM, Wood WI;

WFI: 1985-277976/45.

XX N-PSDB; AAN50054.

PT New recombinant functional human factor VIII or deriv. - useful for treating haemophilia and obtd. as pure prod. by recombinant DNA technology.

XX Disclosure; Fig 10a(10-10C(III)); 109pp; English.

PS The sequence is that of human factor VIII. Amino acids 1-19 are the predicted signal peptide, and amino acids 1-2332 are the predicted mature protein. The protein is produced in pure form and in useful amts. using recombinant DNA techniques. Factor VIII can be used to correct factor VIII deficient plasma, and activates factor X to Xa in the presence of factor IXa, Ca and phospholipid. These to activities are inactivated by antibodies specific for factor VIII. The activity of the prod. is bound to an immobilised monoclonal antibody specific for factor VIII. Factor VIII activity is activated by thrombin. The activity is bound to, and can be eluted from, immobilised von Willebrand factor. Dose of factor VIII is 20-40 units/kg over 8 hours i.v. for maintenance therapy for haemophilia, 40 units/kg for preoperative conditions, or 15-20 units/kg for minor haemorrhaging

Hit wild type
Factor VIII

XX	SQ	Sequence	2351 AA;	Query Match	100.0%; Score 12410; DB 1; Length 2351;	
				Best Local Similarity	100.0%; Pred. No. 0;	
				Matches 2350; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
QY	1	MQIELSTCFELCLRPFSATRRYILGAVELSWDYNQSDIGELPVDARPPRRVKSFPFN	60			
DB	1	MQIELSTCFELCLRPFSATRRYILGAVELSWDYNQSDIGELPVDARPPRRVKSFPFN	60			
QY	61	TSVVYKTLFVEFTDHLFNIAKRPWMGLLPTCAEYVDTWVTLKNMASHPVSLHAV	120			
DB	61	TSVVYKTLFVEFTDHLFNIAKRPWMGLLPTCAEYVDTWVTLKNMASHPVSLHAV	120			
QY	121	GVSYWKASGAEYDDQTSQREKEDKVPFGGSHYVWQVLKNGPMASDPLCLTYSLSH	180			
DB	121	GVSYWKASGAEYDDQTSQREKEDKVPFGGSHYVWQVLKNGPMASDPLCLTYSLSH	180			
QY	181	VDLVKDLNSGLIGALLVCEGSLAKEKTOTLAKTILLFAVDEGKSWSETKNSLMQDRD	240			
DB	181	VDLVKDLNSGLIGALLVCEGSLAKEKTOTLAKTILLFAVDEGKSWSETKNSLMQDRD	240			
QY	241	AASAAWPKQHTWGVNRSPLGICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNH	300			
DB	241	AASAAWPKQHTWGVNRSPLGICHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNH	300			
QY	301	ROASLEISPIITLTAQTLMDLQFLLSCHISSHQHDGMEAYVKVDSCEPRLRMKNNE	360			
DB	301	ROASLEISPIITLTAQTLMDLQFLLSCHISSHQHDGMEAYVKVDSCEPRLRMKNNE	360			
QY	361	EAEYDDDLTDSMDVVRDNDNSPFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVLA	420			
DB	361	EAEYDDDLTDSMDVVRDNDNSPFIQIRSVAKKHPTWVHYIAAEEDWDYAPLVLA	420			
QY	421	PDORSYKSYLNNCPQIRGRKYKVFMAYTDFXTREAIQHEGILGLAYGEVGDTL	480			
DB	421	PDORSYKSYLNNCPQIRGRKYKVFMAYTDFXTREAIQHEGILGLAYGEVGDTL	480			
QY	481	LIIFKQAGRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPFPLPGBIYKYKVTIVEDGP	540			
DB	481	LIIFKQAGRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPFPLPGBIYKYKVTIVEDGP	540			
QY	541	TKSDPCLTRYSSVFNMRDLASLIGPLLI CYKESVDORGNOIMSDKENVILFSVFDE	600			
DB	541	TKSDPCLTRYSSVFNMRDLASLIGPLLI CYKESVDORGNOIMSDKENVILFSVFDE	600			
QY	601	NRSWYLTENIQRLPNPAGVQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWYLLS	660			
DB	601	NRSWYLTENIQRLPNPAGVQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYWYLLS	660			
QY	661	IGAQTDFLSVFFSGYTFKHKWVYEDTLTLPFSGETVPMSENPGMLWILCHNSDFNRG	720			
DB	661	IGAQTDFLSVFFSGYTFKHKWVYEDTLTLPFSGETVPMSENPGMLWILCHNSDFNRG	720			
QY	721	MTALLKVSSCDKNTGDYVEDSYEDI SAILLSKNAIEPRFSQNSRHPSTROKQFNATTI	780			
DB	721	MTALLKVSSCDKNTGDYVEDSYEDI SAILLSKNAIEPRFSQNSRHPSTROKQFNATTI	780			
QY	781	PENDIEKTDWPFAHRTMPKIQNVSSDMLMLRQSPTPHGLSLSDLOEAKYTFSDDDPS	840			
DB	781	PENDIEKTDWPFAHRTMPKIQNVSSDMLMLRQSPTPHGLSLSDLOEAKYTFSDDDPS	840			
QY	841	PGAIDSNNSLSEMTFHPQLHSGDQVFTPESGLOLBNELGTTAATELKKLDPKVSSST	900			
DB	841	PGAIDSNNSLSEMTFHPQLHSGDQVFTPESGLOLBNELGTTAATELKKLDPKVSSST	900			
QY	901	SNNLI STIPSNLAAGTNDTSIGPPMPVHYDSQDITLLFGKKSSPLTBSGGPLSSEE	960			
DB	901	SNNLI STIPSNLAAGTNDTSIGPPMPVHYDSQDITLLFGKKSSPLTBSGGPLSSEE	960			
QY	961	NDSKLLSEGLMNSQESSWGKNVSTSGRLFKGKRAHGPALLTKDNALFKVISTLLKTN	1020			

DB	961	NDSKLLSEGLMNSQESSWGKNVSTSGRLFKGKRAHGPALLTKDNALFKVISTLLKTN	1020
QY	1021	KTSNNSATNRKTHLDGSSILLIENSPSVQNTLESDTEFKVTPLIHDMMLDKNATALRL	1080
DB	1021	KTSNNSATNRKTHLDGSSILLIENSPSVQNTLESDTEFKVTPLIHDMMLDKNATALRL	1080
QY	1081	NHMSNKTTSSKNMVMVQOKKEGPIDPAQNPDMSFFXKMLFLPESARMIQRTGHKNSLNSG	1140
DB	1081	NHMSNKTTSSKNMVMVQOKKEGPIDPAQNPDMSFFXKMLFLPESARMIQRTGHKNSLNSG	1140
QY	1141	QGPSPKOLVSLGPKSVEGQNFLEKKNVUVKGBEFTKOVGLKEMVFPSSNLFETWLDN	1200
DB	1141	QGPSPKOLVSLGPKSVEGQNFLEKKNVUVKGBEFTKOVGLKEMVFPSSNLFETWLDN	1200
QY	1201	LHENNTNHOEKIOEIEBKETTLIQENVVLPOIHTVTGTQNFMKNLFLLSRQNVGSGYD	1260
DB	1201	LHENNTNHOEKIOEIEBKETTLIQENVVLPOIHTVTGTQNFMKNLFLLSRQNVGSGYD	1260
QY	1261	CAYAPVLQDPRSLNDSTNRKTAHFSKKGEEENLEGLGNQTKQIVKRYACTTRISNT	1320
DB	1261	CAYAPVLQDPRSLNDSTNRKTAHFSKKGEEENLEGLGNQTKQIVKRYACTTRISNT	1320
QY	1321	SOQNVTORSKRALQFRLPLEETLEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE	1380
DB	1321	SOQNVTORSKRALQFRLPLEETLEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE	1380
QY	1381	KGAIQTOSPLSDCLTRSHSIPQANRSPPIAKVSSPSPRIPIYLTPLVLPQDNSSHLPAASY	1440
DB	1381	KGAIQTOSPLSDCLTRSHSIPQANRSPPIAKVSSPSPRIPIYLTPLVLPQDNSSHLPAASY	1440
QY	1441	RKDSGVQESSHFLQAGAKNNLSLAILTLEMTGDQREVSGLSGTSATNSVTYKKVENTVLP	1500
DB	1441	RKDSGVQESSHFLQAGAKNNLSLAILTLEMTGDQREVSGLSGTSATNSVTYKKVENTVLP	1500
QY	1501	KDPLPKTSKGVLELPAKHLYQKDLFPTTSNGSPCHLDLVEGSLQGTGEGALKMNEANRP	1560
DB	1501	KDPLPKTSKGVLELPAKHLYQKDLFPTTSNGSPCHLDLVEGSLQGTGEGALKMNEANRP	1560
QY	1561	GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTILSL	1620
DB	1561	GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTILSL	1620
QY	1621	NACESNHAAINEGONKPEIEVWAKQGTRELCSQNPVLPKHQREITRITLOSQDQE	1680
DB	1621	NACESNHAAINEGONKPEIEVWAKQGTRELCSQNPVLPKHQREITRITLOSQDQE	1680
QY	1681	IDYDDTISVEMKKEPDIYDEDEBNSPFSOKTRHYPTAAVERLWDYGMSSSPHVLNR	1740
DB	1681	IDYDDTISVEMKKEPDIYDEDEBNSPFSOKTRHYPTAAVERLWDYGMSSSPHVLNR	1740
QY	1741	AQSGSVQPKKVVQFQFTDGSFTQPLRGELNEHGLGPGYIRAEVEDNIMVTFRNQASR	1800
DB	1741	AQSGSVQPKKVVQFQFTDGSFTQPLRGELNEHGLGPGYIRAEVEDNIMVTFRNQASR	1800
QY	1801	PYSFYSSLISYEDQOQABPRKNFVKNETKTYFWKVQHMAPTKDEDTCKAWAYFSDV	1860
DB	1801	PYSFYSSLISYEDQOQABPRKNFVKNETKTYFWKVQHMAPTKDEDTCKAWAYFSDV	1860
QY	1861	DLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFTTIDETKSWPTENKERNCR	1920
DB	1861	DLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFTTIDETKSWPTENKERNCR	1920
QY	1921	PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVVAQDQRIWYLLSGMSENHISIFSGH	1980
DB	1921	PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVVAQDQRIWYLLSGMSENHISIFSGH	1980
QY	1981	VFTVRKKEEYKALYNLYPGVFTVEMLPKSKAGIWRVBCLIQHLHAGSTLPLVYSNKC	2040
DB	1981	VFTVRKKEEYKALYNLYPGVFTVEMLPKSKAGIWRVBCLIQHLHAGSTLPLVYSNKC	2040
QY	2041	QTPLGWASGHIRDFQITAGQYQWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMII	2100
DB	2041	QTPLGWASGHIRDFQITAGQYQWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMII	2100

CC that ultimately converts soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A deficiency in factor VIII:C is responsible for haemophilia A, which is an X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is activated by plasma proteases, such as thrombin. During activation the mature polypeptide is cleaved to generate heavy and light chain fragments that are further cleaved. The DNA encoding this sequence is mutated, using mutagenic primers, to produce the active Factor VIII:C analogues of the invention (such as AAW11330). The analogues comprise a native Factor VIII:C polypeptide modified at a site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is created. Complexes of two or more of the analogues, nucleic acids and vectors encoding them may be used alone or in conjunction with each other, for the prevention or treatment of active Factor VIII:C deficiency in a mammal. The analogues may be used as immunogens to raise antibodies, and in the treatment of haemophilias, by improvement of haemostasis. The analogues are resistant to proteolytic cleavage and display increased plasma half-life. They may be administered at lower dosages and by different modes of administration

XX
SQ Sequence 2351 AA;

Query Match 100.0%; Score 12410; DB 2; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQIELSTCFELCLLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60
DB 1 MQIELSTCFELCLLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVTILKNWASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVTILKNWASHPVSLHAV 120

QY 121 GVSWKASGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSWKASGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFAVFEDECKSHSEFTKSLMQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFAVFEDECKSHSEFTKSLMQDRD 240

QY 241 AASARAPKMTYNGYVNRSLPGILGCHRKSVYWHVIGMTTPVHSIFLEGHTFLVRNH 300
DB 241 AASARAPKMTYNGYVNRSLPGILGCHRKSVYWHVIGMTTPVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPPQRMKNNE 360
DB 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPPQRMKNNE 360

QY 361 EADYDDDLTDSMDVVRFDNDSFPIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSMDVVRFDNDSFPIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420

QY 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480
DB 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480

QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLIICYKESVDQRCNOJMSDKRNVLFSVPE 600
DB 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLIICYKESVDQRCNOJMSDKRNVLFSVPE 600

QY 601 NESWYLTENIQRFLEPNAGVQLEDPEFQASNMHMSINGYVVDLSQLSVCLHEVAYWYILS 660
DB 601 NESWYLTENIQRFLEPNAGVQLEDPEFQASNMHMSINGYVVDLSQLSVCLHEVAYWYILS 660

QY 661 IGAQTDFLSVFSGYTFKHKWYVEDTLTLFPFSGETVFMSPNPGIWLGCNHSDFRNRG 720
DB 661 IGAQTDFLSVFSGYTFKHKWYVEDTLTLFPFSGETVFMSPNPGIWLGCNHSDFRNRG 720

CC that ultimately converts soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A deficiency in factor VIII:C is responsible for haemophilia A, which is an X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is activated by plasma proteases, such as thrombin. During activation the mature polypeptide is cleaved to generate heavy and light chain fragments that are further cleaved. The DNA encoding this sequence is mutated, using mutagenic primers, to produce the active Factor VIII:C analogues of the invention (such as AAW11330). The analogues comprise a native Factor VIII:C polypeptide modified at a site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is created. Complexes of two or more of the analogues, nucleic acids and vectors encoding them may be used alone or in conjunction with each other, for the prevention or treatment of active Factor VIII:C deficiency in a mammal. The analogues may be used as immunogens to raise antibodies, and in the treatment of haemophilias, by improvement of haemostasis. The analogues are resistant to proteolytic cleavage and display increased plasma half-life. They may be administered at lower dosages and by different modes of administration

XX
SQ Sequence 2351 AA;

Query Match 100.0%; Score 12410; DB 2; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQIELSTCFELCLLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60
DB 1 MQIELSTCFELCLLRFCSATRYIYGAVLSWDYMQSDLGELPVDARPPRPVKSPFN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVTILKNWASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLGPTIQAEVYDVTVTILKNWASHPVSLHAV 120

QY 121 GVSWKASGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSWKASGAEYDDQTSQREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFAVFEDECKSHSEFTKSLMQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTOILHFKIILFAVFEDECKSHSEFTKSLMQDRD 240

QY 241 AASARAPKMTYNGYVNRSLPGILGCHRKSVYWHVIGMTTPVHSIFLEGHTFLVRNH 300
DB 241 AASARAPKMTYNGYVNRSLPGILGCHRKSVYWHVIGMTTPVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPPQRMKNNE 360
DB 301 ROASLEISPTITFLTAQTLLMDLGOFLFCHISSHQHDGMEAYKVVDSCPEPPQRMKNNE 360

QY 361 EADYDDDLTDSMDVVRFDNDSFPIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSMDVVRFDNDSFPIQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420

QY 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480
DB 421 PDORSYKSOYLNNNGPQIRGRYKVRFMAYTDETEKTRAIQHESGILGPLLYGEVDTL 480

QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLIICYKESVDQRCNOJMSDKRNVLFSVPE 600
DB 541 TKSDPRCLTRYYSFVNMRDLASGLIGPLLIICYKESVDQRCNOJMSDKRNVLFSVPE 600

QY 601 NESWYLTENIQRFLEPNAGVQLEDPEFQASNMHMSINGYVVDLSQLSVCLHEVAYWYILS 660
DB 601 NESWYLTENIQRFLEPNAGVQLEDPEFQASNMHMSINGYVVDLSQLSVCLHEVAYWYILS 660

QY 661 IGAQTDFLSVFSGYTFKHKWYVEDTLTLFPFSGETVFMSPNPGIWLGCNHSDFRNRG 720
DB 661 IGAQTDFLSVFSGYTFKHKWYVEDTLTLFPFSGETVFMSPNPGIWLGCNHSDFRNRG 720

721 MTALLKSSCDKNTGDDYEDSYEDISAYLLSKXNAIEPRFSQNSRHSTPSTQKQFNATTI 780
721 MTALLKSSCDKNTGDDYEDSYEDISAYLLSKXNAIEPRFSQNSRHSTPSTQKQFNATTI 780
781 PENDIEKTDWFAHRTMPKIQNVSSDLMLDQSPTPHGLSLSDIQEAKYETFSDDPS 840
781 PENDIEKTDWFAHRTMPKIQNVSSDLMLDQSPTPHGLSLSDIQEAKYETFSDDPS 840
841 PGAIDSNNSISEMTHFRPOLHISGDMVFTBESGLQRLNEKLGTAATELKKLDFKVSST 900
841 PGAIDSNNSISEMTHFRPOLHISGDMVFTBESGLQRLNEKLGTAATELKKLDFKVSST 900
901 SNNLISITPSNLAAGTNDTSSIGPPSNVHYDQLODTLLPGKSSPLTESGGPLSLEE 960
901 SNNLISITPSNLAAGTNDTSSIGPPSNVHYDQLODTLLPGKSSPLTESGGPLSLEE 960
961 NNDKLLLESGLMNSQESSWGKNVSTESGRFLFKGRAGPALLTKDNALFKVSISSLKTN 1020
961 NNDKLLLESGLMNSQESSWGKNVSTESGRFLFKGRAGPALLTKDNALFKVSISSLKTN 1020
1021 KTSNNSATNRKTHIDGSPILLIENSPVMONILSDTEPKVTPLIHDRMLMDKNATLRL 1080
1021 KTSNNSATNRKTHIDGSPILLIENSPVMONILSDTEPKVTPLIHDRMLMDKNATLRL 1080
1081 NMGNKTTSSKNMVMQCKGZGPIPPDAQNPDMSFFKMLFLPESARWTQRTGKNSLNSG 1140
1081 NMGNKTTSSKNMVMQCKGZGPIPPDAQNPDMSFFKMLFLPESARWTQRTGKNSLNSG 1140
1141 QGSPKQVLSIGPEKSVBQGNFLSKNVVKGGEFTKQVGLKEMVFPSSRNLFUTNLDN 1200
1141 QGSPKQVLSIGPEKSVBQGNFLSKNVVKGGEFTKQVGLKEMVFPSSRNLFUTNLDN 1200
1201 LHENNTNHOEKIIOBEIEKKETLIGENVVLPOIHTVTGTHKPMKNLFLLSRQNVGSGYD 1260
1201 LHENNTNHOEKIIOBEIEKKETLIGENVVLPOIHTVTGTHKPMKNLFLLSRQNVGSGYD 1260
1261 GAYAVPLQDFRSLNDSNRTKGTAFHFKKGBEENLEGLGNQTKQIVKCYACTTRISPT 1320
1261 GAYAVPLQDFRSLNDSNRTKGTAFHFKKGBEENLEGLGNQTKQIVKCYACTTRISPT 1320
1321 SQQNVFVQSRKHALQKFLPLBETLEKRIIVDDTSTQMSKNKXHLTPSTLTQIDYNEKE 1380
1321 SQQNVFVQSRKHALQKFLPLBETLEKRIIVDDTSTQMSKNKXHLTPSTLTQIDYNEKE 1380
1381 KGAIQTQSLSDCLTSHSHIPOANRPLPIAKVSSPSPIPIYLVLPQDSSSHLPAASY 1440
1381 KGAIQTQSLSDCLTSHSHIPOANRPLPIAKVSSPSPIPIYLVLPQDSSSHLPAASY 1440
1441 RKKDSGVQSSSHFLOGAKKNLSLAILTLEMTGDQREVGLGTSAINSVTYKKVENTVLP 1500
1441 RKKDSGVQSSSHFLOGAKKNLSLAILTLEMTGDQREVGLGTSAINSVTYKKVENTVLP 1500
1501 KPDLPKTSKVELLPKVHIYQKDLPPFTETSNQSPGHLDLVEGSLGCTGAIKXNEANRP 1560
1501 KPDLPKTSKVELLPKVHIYQKDLPPFTETSNQSPGHLDLVEGSLGCTGAIKXNEANRP 1560
1561 GXVFFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQKSEKPAFKKOTILSL 1620
1561 GXVFFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQKSEKPAFKKOTILSL 1620
1621 NACESNEAIAAINEGNQKPEIEVWAKQGRTERLCSQNPVLRKHORETRITLQSDQBE 1680
1621 NACESNEAIAAINEGNQKPEIEVWAKQGRTERLCSQNPVLRKHORETRITLQSDQBE 1680
1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSQKTRHYFFIAAVERLWDYGMSSSPHVRNR 1740
1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSQKTRHYFFIAAVERLWDYGMSSSPHVRNR 1740
1741 AQSGSVPOPKVQVQFTDGSFTQPLVYRGLNHLGLLGPYIRAEVEDNIMVTRQASR 1800
1741 AQSGSVPOPKVQVQFTDGSFTQPLVYRGLNHLGLLGPYIRAEVEDNIMVTRQASR 1800
1801 PYSFYSSLISYEEDQRCQAEPRKNFVKPNETKTYFWKVQHMAPTKDBFCKAWAYFSDV 1860

1801 PYSFYSSLISYEEDQRCQAEPRKNFVKPNETKTYFWKVQHMAPTKDBFCKAWAYFSDV 1860
1861 DLEKDVHSGLIGLIVCHTNTLNPAGROVTVQEFALFFTFIDETKSWYETENWNCRA 1920
1861 DLEKDVHSGLIGLIVCHTNTLNPAGROVTVQEFALFFTFIDETKSWYETENWNCRA 1920
1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQRIKWYLLSMGSENENIHSIFSGH 1980
1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQRIKWYLLSMGSENENIHSIFSGH 1980
1981 VFTVRKKEBYKVALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEBYKVALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTPILGMASGHIRDFQITASQYQGWAPKARLHYSGSINAWSTKEPPSWIKVDLLAPMI 2100
2041 QTPILGMASGHIRDFQITASQYQGWAPKARLHYSGSINAWSTKEPPSWIKVDLLAPMI 2100
2101 HGKIQGARQKFSLSYISQFIIMVSLDGKKQVTVYRGNSTGTLMVFFGNVDSGSIKNIEN 2160
2101 HGKIQGARQKFSLSYISQFIIMVSLDGKKQVTVYRGNSTGTLMVFFGNVDSGSIKNIEN 2160
2161 PPIIARIYIRLHPHYSTIRLMELMCGDLNSCMPLGMSKASDAQITASSYFTNWEA 2220
2161 PPIIARIYIRLHPHYSTIRLMELMCGDLNSCMPLGMSKASDAQITASSYFTNWEA 2220
2221 TWSFSKARLHQGRSNARVQNNPKEWLQVDFOKTMKVGTGVTQGVKSLTSMYVKEFL 2280
2221 TWSFSKARLHQGRSNARVQNNPKEWLQVDFOKTMKVGTGVTQGVKSLTSMYVKEFL 2280
2281 ISSSQDGHQWTLTFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLIRHPQSWVHQIALRM 2340
2281 ISSSQDGHQWTLTFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLIRHPQSWVHQIALRM 2340
2341 EVLGCBAQDLY 2351
2341 EVLGCBAQDLY 2351
RESULT 3
AAW46245
ID AAW46245 standard; protein; 2351 AA.
XX
AC AAW46245;
XX
DT 25-MAR-2003 (revised)
DT 06-AUG-1998 (first entry)
XX
Human factor III protein sequence.
XX
Replication defective; recombinant retrovirus; RVV; therapeutic protein;
KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
KW hepaticis; thalassemia; phenylketonuria; Lesch-Nyhan syndrome; diabetes;
KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia;
KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
KW Guacher's syndrome; high blood pressure; Alzheimer's disease, factor III;
KW autoimmune; inflammatory disease.
XX
Homo sapiens.
XX
WO9800541-A2.
XX
08-JAN-1998.
XX
02-JUL-1997; 97WO-US011784.
XX
03-JUL-1996; 96US-00645601.
PR 13-AUG-1996; 96US-00696381.
PR 04-JUN-1997; 97US-00869309.
XX
(CHIR) CHIRON CORP.
XX

PI Jolly DJ, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M;
 PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;
 PI Mittelstaedt DM, Prussak CE, Greengard J, Lee R;

XX WPI: 1998-086966/08.
 DR N-PSDB; AAV19580.

XX New replication defective recombinant retro-viruses - which can be
 PT administered to provide long term systemic expression of therapeutic
 PT protein in blood, useful in, e.g. treating hyper-coagulable disorders.
 XX
 XX Example 28; Page 203-210; 272pp; English.

XX This is the human factor III sequence. The encoding DNA is used in the
 CC construction of recombinant retroviral vectors expressing human factor
 CC viii. The specification provides the preparation of replication defective
 CC recombinant retrovirus (RRV) expressing a therapeutic protein. The RRV
 CC preparation is resistant to degradation by human complement and is
 CC capable of inducing long term systemic expression of the therapeutic
 CC protein when administered intravenously to a human. The long term
 CC systemic expression results in a measurable level of the therapeutic
 CC protein being produced in the blood of the human for a period of at least
 CC 30 days after the administration of the RRV vector preparation. RRV's can
 CC be used for in vivo delivery of therapeutic protein to treat, e.g.
 CC haemophilia A, haemophilia B, thrombosis, hypercoagulable disorders,
 CC liver diseases such as hepatitis, disorders such as thalassemia,
 CC phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency
 CC (SCID), cystic fibrosis, Duchenne's Muscular Dystrophy, inherited
 CC emphysema, familial hypercholesterolemia, diabetes, hypopituitarism,
 CC ageneine deaminase deficiency, alapha-antitrypsin deficiency, Gaucher's
 CC syndrome, anaemia, infections such as HIV infection, high blood pressure,
 CC Alzheimer's disease, autoimmune or inflammatory disease or graft versus
 CC host disease. RRV's are capable of surviving inactivation in human serum
 CC thereby allowing efficient gene transfer over prolonged periods of time.
 XX (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 2351 AA;

Query Match 100.0%; Score 12410; DB 2; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOELSTCFELCLLRFCFSATRRYVLGAVESWDMQSDGLGELPVDARPPRPVPSFFPN 60
 DB 1 MOELSTCFELCLLRFCFSATRRYVLGAVESWDMQSDGLGELPVDARPPRPVPSFFPN 60

QY 61 TSVYKKTLEVEFTDHLFNIAKRPZPMGLLQFTIOAEVYDVTWITLKNMASHPVSLHAV 120
 DB 61 TSVYKKTLEVEFTDHLFNIAKRPZPMGLLQFTIOAEVYDVTWITLKNMASHPVSLHAV 120

QY 121 GVSYWKASGAGAYDDOTSOREKEDDKVFGSGHTYVQVLKENGPMASDPLCLTYSYLSH 180
 DB 121 GVSYWKASGAGAYDDOTSOREKEDDKVFGSGHTYVQVLKENGPMASDPLCLTYSYLSH 180

QY 181 VDLVLDNSGLIGALLVRCGSLAKEKTLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240
 DB 181 VDLVLDNSGLIGALLVRCGSLAKEKTLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPQHTVNGVYVNSRLPGLIGCHRSKVYVHVGMTTPEVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPQHTVNGVYVNSRLPGLIGCHRSKVYVHVGMTTPEVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISITITLTAQTLMDLGOFLLSCHISSHQHDGMEAYVKVSDSCPEEPQLRMKNE 360
 DB 301 ROASLEISITITLTAQTLMDLGOFLLSCHISSHQHDGMEAYVKVSDSCPEEPQLRMKNE 360

QY 361 EAEDYDDDLTDSMDVVRDSDNSPSFIOIRSVAKKHPTWVHYIAAEBEDMDYAPLVLA 420
 DB 361 EAEDYDDDLTDSMDVVRDSDNSPSFIOIRSVAKKHPTWVHYIAAEBEDMDYAPLVLA 420

QY 421 PDRSYKSYQLNNGPQIRGRKYKVFPMAYTDTFTKTREAIQHESGILGPLLYGEVGDIL 480
 DB 421 PDRSYKSYQLNNGPQIRGRKYKVFPMAYTDTFTKTREAIQHESGILGPLLYGEVGDIL 480

QY 481 LLIIFKQASRPYNIYHGIITDVRLPYLSRRIPKGVKHLKDPFIPILGGEIFKFKWTVTVDGP 540
 DB 481 LLIIFKQASRPYNIYHGIITDVRLPYLSRRIPKGVKHLKDPFIPILGGEIFKFKWTVTVDGP 540

QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGLIPLLICYKESVDQGNQIMSDKRNVLFSVFDE 600
 DB 541 TKSDPRCLTRYSSFVNMRDLASGLIGLIPLLICYKESVDQGNQIMSDKRNVLFSVFDE 600

QY 601 NRSWYLNTQREFLNPNAGVQLEDPEQASNMHSINGYVFDLSQLSVCLHEVAYWILS 660
 DB 601 NRSWYLNTQREFLNPNAGVQLEDPEQASNMHSINGYVFDLSQLSVCLHEVAYWILS 660

QY 661 IGAQTDLSLFFSYGTFKHWYEDTTLTPPFSGETVFMGMENPGLWILGCHNSDFNRG 720
 DB 661 IGAQTDLSLFFSYGTFKHWYEDTTLTPPFSGETVFMGMENPGLWILGCHNSDFNRG 720

QY 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTKQFNATTI 780
 DB 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTKQFNATTI 780

QY 781 PENDIEKTDWFAHRTMPKTIQNVSSDLIMLLRQSPTHGLISLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTDWFAHRTMPKTIQNVSSDLIMLLRQSPTHGLISLSDLOEAKYETFSDDPS 840

QY 841 PGALDSNNSISEMTHFRPQLHHSQDMVTTPESGLQLRINEKLGTTAATLKKLDFKVSST 900
 DB 841 PGALDSNNSISEMTHFRPQLHHSQDMVTTPESGLQLRINEKLGTTAATLKKLDFKVSST 900

QY 901 SNNLITIPSDNLAAGTNTSSLGPPMPVHYDSOLDTTLFGKSSPLTESGGPLSLSEE 960
 DB 901 SNNLITIPSDNLAAGTNTSSLGPPMPVHYDSOLDTTLFGKSSPLTESGGPLSLSEE 960

QY 961 NNDKLLLESGLMNSQESSWGKNVASTESGRVFKGRAGPALLTKDNALFKVSISLLKTN 1020
 DB 961 NNDKLLLESGLMNSQESSWGKNVASTESGRVFKGRAGPALLTKDNALFKVSISLLKTN 1020

QY 1021 KTSNKSATNRKTHIDGFSLLIENSFVWQNILEDTEFEKVTPLIHRLMKDKNATALRL 1080
 DB 1021 KTSNKSATNRKTHIDGFSLLIENSFVWQNILEDTEFEKVTPLIHRLMKDKNATALRL 1080

QY 1081 NHMKNKTTSSKNMVMVQKKEGPIPPDAQNDDMGFFKMLFIPESARWQORTHKNSLNSG 1140
 DB 1081 NHMKNKTTSSKNMVMVQKKEGPIPPDAQNDDMGFFKMLFIPESARWQORTHKNSLNSG 1140

QY 1141 QGSPKQVLSGLPEKSVGQNFLEKKNVVGKGFETKDVGLKENVFPSSRNLPFLTNLDN 1200
 DB 1141 QGSPKQVLSGLPEKSVGQNFLEKKNVVGKGFETKDVGLKENVFPSSRNLPFLTNLDN 1200

QY 1201 LHENNTNQEKKIQEIEIEKKTETLIQENVVLPQIHVTVTCTKNFMKNLFLLLSTRQNVESYD 1260
 DB 1201 LHENNTNQEKKIQEIEIEKKTETLIQENVVLPQIHVTVTCTKNFMKNLFLLLSTRQNVESYD 1260

QY 1261 GAYAVLPQDPSLNDSTNRKHTAHFSPKGEENLEGNGQTKQIVKCYACTTISPN 1320
 DB 1261 GAYAVLPQDPSLNDSTNRKHTAHFSPKGEENLEGNGQTKQIVKCYACTTISPN 1320

QY 1321 SOONFVTOQRSKRALQKPLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SOONFVTOQRSKRALQKPLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

QY 1381 KGAIQTQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPFSIRPIYLRVLFDQDNSSHPAASY 1440
 DB 1381 KGAIQTQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPFSIRPIYLRVLFDQDNSSHPAASY 1440

QY 1441 RKOSGVQESSHFLGAKAKNNLSAILTLEMTGDQREVGLSGTSATNSVTYKKVENTVLP 1500
 DB 1441 RKOSGVQESSHFLGAKAKNNLSAILTLEMTGDQREVGLSGTSATNSVTYKKVENTVLP 1500

QY 1501 KPDLPKTSKVELLPKVHIYQKDLPTFTSNGSPGHLDLVEGSLLOGTEGALKWNEANRP 1560
 DB 1501 KPDLPKTSKVELLPKVHIYQKDLPTFTSNGSPGHLDLVEGSLLOGTEGALKWNEANRP 1560

```

QY 1561 GKVPFLVATSSAKTBSKLLDPLAWNHYGTQIPKEWKSQBSKPEKTAFFKXDTILSL 1620
DB 1561 GKVPFLVATSSAKTBSKLLDPLAWNHYGTQIPKEWKSQBSKPEKTAFFKXDTILSL 1620
QY 1621 NACESNHAIAAINBGNKPELEVWAKQGRTERLCSQNPVYLKHPHREITRTTLQSDEE 1680
DB 1621 NACESNHAIAAINBGNKPELEVWAKQGRTERLCSQNPVYLKHPHREITRTTLQSDEE 1680
QY 1681 IDYDDTISVENMKEDFDIYDEENQSPRSFQKTRHYTAAVERLWDYGMSSPHVLNR 1740
DB 1681 IDYDDTISVENMKEDFDIYDEENQSPRSFQKTRHYTAAVERLWDYGMSSPHVLNR 1740
QY 1741 AQSGVPOFKKVVPOEFDGFTQPLVGEINELHGLLGPVIRAEVDENIIVTFNQASR 1800
DB 1741 AQSGVPOFKKVVPOEFDGFTQPLVGEINELHGLLGPVIRAEVDENIIVTFNQASR 1800
QY 1801 PYSFYSSLSIYEEDQOGAEPKRNFKVKNETKYPKWQVHNAFTKDFDCKAWAYFSDV 1860
DB 1801 PYSFYSSLSIYEEDQOGAEPKRNFKVKNETKYPKWQVHNAFTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLIGPLIVCHTNTLNPAHGRQVTVQBEALFTIFDETKSWYFTENMERNCA 1920
DB 1861 DLEKDVHSLIGPLIVCHTNTLNPAHGRQVTVQBEALFTIFDETKSWYFTENMERNCA 1920
QY 1921 PCNIQMEDPTFKENYRFAHNGIYIMDTPLGLVMAQDQIRWYILSMGSGNENIHSIFSGH 1980
DB 1921 PCNIQMEDPTFKENYRFAHNGIYIMDTPLGLVMAQDQIRWYILSMGSGNENIHSIFSGH 1980
QY 1981 VFTVRKKEEYKMALNLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMALNLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QPLGWASGHIRDFOITASQGVQWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMII 2100
DB 2041 QPLGWASGHIRDFOITASQGVQWAPKLARLHYSGSINAWSTKBPFSWIKVDLLAPMII 2100
QY 2101 HGKITQGARQKFSLSIYISQFIIMYSIDGKKWQTYRGNSTGTLWFFGNVDSSGIXHNIPN 2160
DB 2101 HGKITQGARQKFSLSIYISQFIIMYSIDGKKWQTYRGNSTGTLWFFGNVDSSGIXHNIPN 2160
QY 2161 PPIIARYIRLHPTHYSIRTLRMELMGCDLNSCMLPLGWSKASISDAQITASSYFTNMPA 2220
DB 2161 PPIIARYIRLHPTHYSIRTLRMELMGCDLNSCMLPLGWSKASISDAQITASSYFTNMPA 2220
QY 2221 TWSPSKARLHLOGRNANWAPQVNNPKWLQVDFQKTMKVGTGTVQGVKSLLSMVKKEFL 2280
DB 2221 TWSPSKARLHLOGRNANWAPQVNNPKWLQVDFQKTMKVGTGTVQGVKSLLSMVKKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLYRIRIHPQSWVHQIALRM 2340
DB 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLYRIRIHPQSWVHQIALRM 2340
QY 2341 EVLGCFAQDLY 2351
DB 2341 EVLGCFAQDLY 2351

```

RESULT 4

AAW44373

ID AAW44373 standard; protein; 2351 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

FT Domain 711..1694
/note="B-domain"
XX
XX
XX W09800542-A2.
XX
XX 08-JAN-1998.
XX
XX 02-JUL-1997; 97WO-US0111785.
XX
XX 03-JUL-1996; 96US-00645601.
XX 13-AUG-1996; 96US-00696381.
XX 04-JUN-1997; 97US-00869309.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Jolly DJ, Barber JR, Chang SM, Respass JG, Allen JR, Boder M;
XX Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;
XX Mittelstaedt DM, Prussak CE, Greengard J;
XX
XX WPI; 1998-086967/08.
XX N-PSDB; AAV15359.
XX
XX New replication defective recombinant retroviruses - which express B
XX domain-deleted human factor VIII or human factor IX for the treatment of
XX haemophilia.
XX
XX Example 28; Page 164-166; 236pp; English.
XX
XX This polypeptide comprises human Factor VIII. The invention relates to
XX preparations of replication defective recombinant retrovirus (RV)
XX expressing a B domain-deleted human factor VIII protein, such as the SQN
XX deletion mutant (see AAW44372), where the recombinant RV is capable of
XX infecting human cells, is resistant to degradation by human complement
XX and is capable of inducing long-term (at least 30 days and up to 6 months
XX or longer post-injection) systemic expression of factor VIII when
XX administered to a haemophilia A patient
XX
XX Sequence 2351 AA;

```

```

Query Match 100.0%; Score 12410; DB 2; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MQIELSTCFFCLLRFCSATRRYYLGAVELSWDMQSDLGELPVDARPPRVPKSFPPN 60
DB 1 MQIELSTCFFCLLRFCSATRRYYLGAVELSWDMQSDLGELPVDARPPRVPKSFPPN 60
QY 61 TSVVYKKTILFVEPTDHLNIAKPRPPNMGLLGFTTQAEVYDVITLKNMASHPVSLHAV 120
DB 61 TSVVYKKTILFVEPTDHLNIAKPRPPNMGLLGFTTQAEVYDVITLKNMASHPVSLHAV 120
QY 121 GVSVMKASGEGAEYDDQTSOREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLH 180
DB 121 GVSVMKASGEGAEYDDQTSOREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLH 180
QY 181 VDLVXDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQDRD 240
DB 181 VDLVXDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQDRD 240
QY 241 AASARAWPKQHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
DB 241 AASARAWPKQHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLLMDLQGFLLSCHISSHQHDGMEAYVKVDSCPEEPOLAKMNE 360
DB 301 ROASLEISPIITFLTAQTLLMDLQGFLLSCHISSHQHDGMEAYVKVDSCPEEPOLAKMNE 360
QY 361 EAEVDYDDLTDSGMDVVRPDDNNSPSFIOIRVAKKHPKTVWHYIAAEEDWDYAPLVIA 420
DB 361 EAEVDYDDLTDSGMDVVRPDDNNSPSFIOIRVAKKHPKTVWHYIAAEEDWDYAPLVIA 420
QY 421 PDDRSYKSOYLNNGPORIGRKVKYREMAVYTDFTFKTREAIQHOESGILGPLYGEGDTL 480
DB 421 PDDRSYKSOYLNNGPORIGRKVKYREMAVYTDFTFKTREAIQHOESGILGPLYGEGDTL 480

```

Location/Qualifiers

421 PDDSYKSYQVLLNNGPQRIGRKXKVRFMAYTDTEFTKTRAIQHESGILGPLLYGEVGDTL 480
481 LIIPKQASRPYNYPHGIIIDVRPLYSRRILPKGVKHLKDPPIIPGBIFPKYKWTVTVEDGP 540
481 LIIPKQASRPYNYPHGIIIDVRPLYSRRILPKGVKHLKDPPIIPGBIFPKYKWTVTVEDGP 540
541 TKSDPRCLTRYSSVFVNERDLASGLIGPLILICYKESVDORGNOIMSDKRNVLIFSVD 600
541 TKSDPRCLTRYSSVFVNERDLASGLIGPLILICYKESVDORGNOIMSDKRNVLIFSVD 600
601 NRSYLTENIQRFLENPAGVQLEDPEFQASNIMHSINGYVFDLSQVCLHEVAYWILS 660
601 NRSYLTENIQRFLENPAGVQLEDPEFQASNIMHSINGYVFDLSQVCLHEVAYWILS 660
661 IGAOTDELVSFPFGYTFPKHKKVYEDTLTLPSFGETVFMSENPLMTLGCNDSDFNRG 720
661 IGAOTDELVSFPFGYTFPKHKKVYEDTLTLPSFGETVFMSENPLMTLGCNDSDFNRG 720
721 MTALLKVSCKDNTGDYVEDYEDISAYLLSKNNAIEPRSPSONSRHPSRTRQKQFNATTI 780
721 MTALLKVSCKDNTGDYVEDYEDISAYLLSKNNAIEPRSPSONSRHPSRTRQKQFNATTI 780
781 PENDIEKTDPAFAHRTMPKIONVSSSLLMLLQSPTPHGLSLSLDOEAKYETPSDDPS 840
781 PENDIEKTDPAFAHRTMPKIONVSSSLLMLLQSPTPHGLSLSLDOEAKYETPSDDPS 840
841 PGALDSNLSSEYTHFRPQLHHSQDMVTPESGLQLRLNEKLGTTAAETELKGLDFKVSST 900
841 PGALDSNLSSEYTHFRPQLHHSQDMVTPESGLQLRLNEKLGTTAAETELKGLDFKVSST 900
901 SNNLITIPSDNLAAGTDNTSLGPPSPVHYDSQDITTLFGKSSPLTESGGPLSLSEE 960
901 SNNLITIPSDNLAAGTDNTSLGPPSPVHYDSQDITTLFGKSSPLTESGGPLSLSEE 960
961 NNDKSLBESGLMNSQESGKGNVSTESGRLPKGRAGPALLTKDNALFKVSIILLKTN 1020
961 NNDKSLBESGLMNSQESGKGNVSTESGRLPKGRAGPALLTKDNALFKVSIILLKTN 1020
1021 KTSNNSATNRKTHIDQPSLLIENSFVWQNIIESDTEPKKVTPLIHDRMLMDKNATALRL 1080
1021 KTSNNSATNRKTHIDQPSLLIENSFVWQNIIESDTEPKKVTPLIHDRMLMDKNATALRL 1080
1081 NHMSNKTTSKKNMVMQKKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
1081 NHMSNKTTSKKNMVMQKKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
1141 QGFSPKQLVSLGPEKSVEGQNFLESKNVVVGKGEFTKDVGLKEMVFPSSRNLFITNLDN 1200
1141 QGFSPKQLVSLGPEKSVEGQNFLESKNVVVGKGEFTKDVGLKEMVFPSSRNLFITNLDN 1200
1201 LKENNTHOKEKTOBEIEKKEKTELIQSNVLPQTHVTGKNFMKNLFLLSSTRQNVESYD 1260
1201 LKENNTHOKEKTOBEIEKKEKTELIQSNVLPQTHVTGKNFMKNLFLLSSTRQNVESYD 1260
1261 GAYAPVLDQFRSLDSTNRTKHTAHFSKKGEENLEGLGNQTKQIVEKYACTTRISPNPT 1320
1261 GAYAPVLDQFRSLDSTNRTKHTAHFSKKGEENLEGLGNQTKQIVEKYACTTRISPNPT 1320
1321 SQNFVTSKRAKQFRPLBETELKRLIIVDDTSTQWSKNMKHITPSTLTQIDYNEKE 1380
1321 SQNFVTSKRAKQFRPLBETELKRLIIVDDTSTQWSKNMKHITPSTLTQIDYNEKE 1380
1381 KGAITQSPISDCITRSHSIPQANRSLPIAKVSSFPISIRPIYLRVLFQDNSSHLPAASY 1440
1381 KGAITQSPISDCITRSHSIPQANRSLPIAKVSSFPISIRPIYLRVLFQDNSSHLPAASY 1440
1441 RKKDSGVQESSHFLQGAKNKLSLAJLTLEMTGDQREVSLGTSATNSVTYKXVENTVLP 1500
1441 RKKDSGVQESSHFLQGAKNKLSLAJLTLEMTGDQREVSLGTSATNSVTYKXVENTVLP 1500
1501 KPDLPTSGVKEPLLKPVHIIYQKOLFPTETNGSPGHLDLVEGILLQGTGSAIKWNEANRP 1560
1501 KPDLPTSGVKEPLLKPVHIIYQKOLFPTETNGSPGHLDLVEGILLQGTGSAIKWNEANRP 1560

1561 GKYPFLRVATESSAKTPSKLLDPLAWNHYGTQPKBEWKSQBSKPEKTAFFKXOTILSL 1620
1561 GKYPFLRVATESSAKTPSKLLDPLAWNHYGTQPKBEWKSQBSKPEKTAFFKXOTILSL 1620
1621 NACSNHAIAMINEGQNKPEIEVTWAKQGTRELCSQNPVLPKHOREIIRTTLOSQOEE 1680
1621 NACSNHAIAMINEGQNKPEIEVTWAKQGTRELCSQNPVLPKHOREIIRTTLOSQOEE 1680
1681 IDYDDTISVMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGMSSSHVILNR 1740
1681 IDYDDTISVMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGMSSSHVILNR 1740
1741 AQGSVPQFKKVPQEFDTGSGFTQPLRGELNEHLGLLGPVIRAEVDENIMVTFRNOASR 1800
1741 AQGSVPQFKKVPQEFDTGSGFTQPLRGELNEHLGLLGPVIRAEVDENIMVTFRNOASR 1800
1801 PYSYSSLSIYSEEDQROGAPRKNFVKPNETKTVFKVQVHHVAPTKEFDCAKWAYFSDV 1860
1801 PYSYSSLSIYSEEDQROGAPRKNFVKPNETKTVFKVQVHHVAPTKEFDCAKWAYFSDV 1860
1861 DLEKDVHSGILGPLLVCHTNTLNPAHGQVTVQBFALFFTFIDETKSWYFTENNERCRA 1920
1861 DLEKDVHSGILGPLLVCHTNTLNPAHGQVTVQBFALFFTFIDETKSWYFTENNERCRA 1920
1921 PCNIQMEDPTPKENYRFHAINGYIMDTLPGILVMAQDQRIRWYLLSMGSNENIHSIHFSGH 1980
1921 PCNIQMEDPTPKENYRFHAINGYIMDTLPGILVMAQDQRIRWYLLSMGSNENIHSIHFSGH 1980
1981 VFTVRKKEEYKMALYNLYPGYFETVEMLPKAGIWRVECLIGEHLHAGMSTLELVYSNKC 2040
1981 VFTVRKKEEYKMALYNLYPGYFETVEMLPKAGIWRVECLIGEHLHAGMSTLELVYSNKC 2040
2041 QTPLGMSGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
2041 QTPLGMSGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
2101 EGKIQTOGARQFSSLIYISQFIIMYSLDGKKQWQTVRGNSTGTLMVFFGNVDSGKIHENFN 2160
2101 EGKIQTOGARQFSSLIYISQFIIMYSLDGKKQWQTVRGNSTGTLMVFFGNVDSGKIHENFN 2160
2161 PPIIARYIRLHPHYSIESTIRMEIWMELGCDLNSCMLPGWESKAIISDAQITASSYTNMFA 2220
2161 PPIIARYIRLHPHYSIESTIRMEIWMELGCDLNSCMLPGWESKAIISDAQITASSYTNMFA 2220
2221 TWSPSKARLHLQGRNANRPVANNPKWLQVDFOKTMKVTVTTQGVKSLLTSMYVKEFL 2280
2221 TWSPSKARLHLQGRNANRPVANNPKWLQVDFOKTMKVTVTTQGVKSLLTSMYVKEFL 2280
2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSGLDPPLLTRYLRHHPQSWHQAIALRM 2340
2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSGLDPPLLTRYLRHHPQSWHQAIALRM 2340
2341 EVLGCEAQDLY 2351
2341 EVLGCEAQDLY 2351

RESULT 5
AAY52537
ID AAY52537 standard; protein; 2351 AA.

XX AAY52537;

XX AC

XX XX

XX 28-FEB-2000 (first entry)

XX Human full-length factor VIII.

DE Factor VIII; haemophilia; proteolysis; heavy chain; light chain;

XX secretion; von Willebrand Factor; vWF; C2 domain; intron 22; inversion;

XX non-functional; bleeding disorder; coagulation; treatment.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..19 "Signal peptide"
 FT Protein /note= "1708"
 FT Protein /note= "Maximum length human factor VIII heavy chain"
 FT Protein 1709..2351
 FT Protein /note= "Human factor VIII light chain"
 XX WO959622-A1.
 FN 25-NOV-1999.
 XX 17-MAY-1999; 99WO-US010872.
 XX 20-MAY-1998; 98US-00082000.
 XX (GETH) GENENTECH INC.
 PA Vohar GA;
 PI MPI: 2000-053195/04.
 DR N-PSDB; AA238604.
 XX New method for treating hemophilia A.
 PT Disclosure; Fig 9; 50pp; English.
 PS This sequence represents the full-length human factor VIII. In this form,
 CC the protein is 300 kD in size with the domain structure A1-A2-B-A3-C1-C2.
 CC However, prior to secretion, this 300 kD protein is proteolytically cleaved into a
 CC heavy chain (A1-A2-B, with continued proteolysis within the B domain
 CC resulting in molecules of varying length) and a light chain (A3-C1-C2)
 CC that remains non-covalently attached to the heavy chain. Upon secretion,
 CC factor VIII is rapidly cleared from the circulation unless it is bound by
 CC the plasma protein von Willebrand Factor (vWF). Factor VIII binds to vWF
 CC through the light chain, with known binding sites at the N-terminus and
 CC within the C2 domain. Haemophilia A is frequently caused by an intron 22
 CC inversion in the factor VIII gene, which separates the gene into two
 CC sections, exons 1-22 becoming inverted and localised telomeric to the
 CC original site, while exons 23-26 remain isolated at the original site.
 CC Exons 23-26 a portion of the C1 and all of the C2 domains, without which
 CC factor VIII is non-functional. The factor VIII gene product of
 CC individuals with this mutation thus comprises domains A1-A2-B-A3 plus a
 CC fragment of C1, which on proteolysis is non-functional, resulting in a
 CC bleeding disorder. The invention relates to a novel method for for
 CC treating haemophilia A in a mammal able to produce the factor VIII heavy
 CC chain. The method comprises administering to the mammal either an
 CC effective amount of factor VIII light chain, or a gene encoding it, and
 CC may be useful for treating patients such as those whose haemophilia A is
 CC caused by intron 22 inversion. The recombinant factor VIII products of
 CC this invention are derived from well-characterised starting materials
 CC which therefore reduces the risk of pathogenic infection which was
 CC previously a problem when using donated plasma. Furthermore, the
 CC invention provides a more economic and potentially more effective
 CC treatment for haemophilia. There is also a need of providing factor VIII
 CC activity to patients who produce or are at risk of producing antibodies
 CC against full-length factor VIII
 XX Sequence 2351 AA;
 SQ
 Query Match 100.0%; Score 12410; DB 3; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MQIELSTCFELCLLRFCFSATRYVYLGAVELSDWYMSDYGELPVDARRPPRYPKGFPPN 60
 DB 1 XQIELSTCFELCLLRFCFSATRYVYLGAVELSDWYMSDYGELPVDARRPPRYPKGFPPN 60
 QY 61 TSVVYKXTLVEFTDHLFNIAPKRPVWMLGPTIQAEVYDTVITLKNMASHPVSLHAV 120
 DB 61 TSVVYKXTLVEFTDHLFNIAPKRPVWMLGPTIQAEVYDTVITLKNMASHPVSLHAV 120

QY 121 GVSYYKASGAEAYDDQTSOREKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYSYLSH 180
 DB 121 GVSYYKASGAEAYDDQTSOREKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYSYLSH 180
 QY 181 VDLVKDLSNGLIGALLVCRGSLAKKKTQTLHKSTILLFAHPDEGKSWHSETKNSIMQDRD 240
 DB 181 VDLVKDLSNGLIGALLVCRGSLAKKKTQTLHKSTILLFAHPDEGKSWHSETKNSIMQDRD 240
 QY 241 AASARAWPKMTVNGYVNRSLPGLTGCHRKSVYWEVIGMGTTPPEVHSIFLECHTFLVRNH 300
 DB 241 AASARAWPKMTVNGYVNRSLPGLTGCHRKSVYWEVIGMGTTPPEVHSIFLECHTFLVRNH 300
 QY 301 RQASLEISPIITFLTAQTLMLDLGQFLFESCHISSHQHDCNEAYVVDSPCEPQRLMKONE 360
 DB 301 RQASLEISPIITFLTAQTLMLDLGQFLFESCHISSHQHDCNEAYVVDSPCEPQRLMKONE 360
 QY 361 EAEYDDDLTSEMDVVRFDNNSPFIQIRSVAKKHPKTVHYIAAEEEDWDVAPLVLA 420
 DB 361 EAEYDDDLTSEMDVVRFDNNSPFIQIRSVAKKHPKTVHYIAAEEEDWDVAPLVLA 420
 QY 421 PDRSYKSOYLNNGPQRIGRYKVKVRFMAYTDETFKTREAIQHESGILGLPYGVGDTL 480
 DB 421 PDRSYKSOYLNNGPQRIGRYKVKVRFMAYTDETFKTREAIQHESGILGLPYGVGDTL 480
 QY 481 LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKWTIVVEDGP 540
 DB 481 LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIIFYKWTIVVEDGP 540
 QY 541 TKSDPRCLTRYSSVNMERDLASGLIGPLIICYKESVDQGNQIMSKRVILSVDE 600
 DB 541 TKSDPRCLTRYSSVNMERDLASGLIGPLIICYKESVDQGNQIMSKRVILSVDE 600
 QY 601 NRSWYLTENIQRFLEPNPAGVQLEDPEFOASNMHISINGYVFDLSQLSVCLHEVAYWYLS 660
 DB 601 NRSWYLTENIQRFLEPNPAGVQLEDPEFOASNMHISINGYVFDLSQLSVCLHEVAYWYLS 660
 QY 661 IQAQTDFLSVPSGTYFKHKMYEDTLTLPESGETVFMENPGLWILGCHNSDFRNG 720
 DB 661 IQAQTDFLSVPSGTYFKHKMYEDTLTLPESGETVFMENPGLWILGCHNSDFRNG 720
 QY 721 MTALLKVSCKDNTQDYYEDSVEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 DB 721 MTALLKVSCKDNTQDYYEDSVEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTDVPFAHRTMPMKIKNVSSSLLMLLRSPPTPHGLSLSDLOBAKYETTSDDPS 840
 DB 781 PENDIEKTDVPFAHRTMPMKIKNVSSSLLMLLRSPPTPHGLSLSDLOBAKYETTSDDPS 840
 QY 841 PGALDNNLSLSEMTFRPOLHSGDMVTFPESGLQRLNEKLGTTAATLKLDFKVSST 900
 DB 841 PGALDNNLSLSEMTFRPOLHSGDMVTFPESGLQRLNEKLGTTAATLKLDFKVSST 900
 QY 901 SNKLSTIPSDNLAAGTDMTSSLGPPSPVHYDSQDITTLFCKKSPSTESGGPLSLSEE 960
 DB 901 SNKLSTIPSDNLAAGTDMTSSLGPPSPVHYDSQDITTLFCKKSPSTESGGPLSLSEE 960
 QY 961 NNDKLLSEGLMNSQSSGKGVSTESGRLPKGRAGHPALLTKONALFKYSISLTKN 1020
 DB 961 NNDKLLSEGLMNSQSSGKGVSTESGRLPKGRAGHPALLTKONALFKYSISLTKN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWONILESDFPKKVTPLIHDRMLDKNATLRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWONILESDFPKKVTPLIHDRMLDKNATLRL 1080
 QY 1081 NMSNKTSSKNMVMYQKKEGFIIPDAQNPDMSPFKMLFLPESARWIQRTKNSLNSG 1140
 DB 1081 NMSNKTSSKNMVMYQKKEGFIIPDAQNPDMSPFKMLFLPESARWIQRTKNSLNSG 1140
 QY 1141 QGSPKQLVSLGPEKVEGQNFSEKNKVVVGKGETKQVGLKEMVFPSSRLFTNLN 1200
 DB 1141 QGSPKQLVSLGPEKVEGQNFSEKNKVVVGKGETKQVGLKEMVFPSSRLFTNLN 1200
 QY 1201 LHENNTNOKKIQEBIEKKTILQENVVLPPOINTVTGTFKFMKNLFLLLSTRQNVESYD 1260

Db 1201 LHEHNTNQEKIQEELEKZETLLQENVVLQIHVGTGKNFMKNLFLSTRQNVESYD 1260
Qy 1261 GAYAPVLQDFSLNDSTNRTKHTAHFSKKGBEENLEGLNQTQKQIVKQACTTRISPT 1320
Db 1261 GAYAPVLQDFSLNDSTNRTKHTAHFSKKGBEENLEGLNQTQKQIVKQACTTRISPT 1320
Qy 1321 SQQNFTQSRKALQKPLPLEETLEKRIIVDDTSTQMSKNMKHLPSTLTQIDYNEKE 1380
Db 1321 SQQNFTQSRKALQKPLPLEETLEKRIIVDDTSTQMSKNMKHLPSTLTQIDYNEKE 1380
Qy 1381 KGAITQSPSLCLTRSHSIPQANSRPIAKVSPSPSIRPIYLTRVLFPQDNSSHLPAASY 1440
Db 1381 KGAITQSPSLCLTRSHSIPQANSRPIAKVSPSPSIRPIYLTRVLFPQDNSSHLPAASY 1440
Qy 1441 RKDGGVOEESHFTQGAKKNNLSAILTLEMTGQOREVSGISGTATNSVTYKKVENIVLP 1500
Db 1441 RKDGGVOEESHFTQGAKKNNLSAILTLEMTGQOREVSGISGTATNSVTYKKVENIVLP 1500
Qy 1501 KPDLPKTSCKVELLPKVHIYOKDLFPPTSTNGSPGHLDLVEGSLLOQTGEGAIKNEANRP 1560
Db 1501 KPDLPKTSCKVELLPKVHIYOKDLFPPTSTNGSPGHLDLVEGSLLOQTGEGAIKNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKQESKPEKTAFFKKDTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKQESKPEKTAFFKKDTILSL 1620
Qy 1621 NACENHAIAINEQNKPEIEVTWAKQGRTERLCSQNPVULKRHQRBITRTTLOSQDEE 1680
Db 1621 NACENHAIAINEQNKPEIEVTWAKQGRTERLCSQNPVULKRHQRBITRTTLOSQDEE 1680
Qy 1681 IDYDDTISVEMKEDFDYDEENOSPSFOKTRHYFIAAVERLWDYGMSSPHVLNRR 1740
Db 1681 IDYDDTISVEMKEDFDYDEENOSPSFOKTRHYFIAAVERLWDYGMSSPHVLNRR 1740
Qy 1741 AQSGSVPOFKKVVFQFTDGSFTQPLRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASR 1800
Db 1741 AQSGSVPOFKKVVFQFTDGSFTQPLRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASR 1800
Qy 1801 PYSFYSSLIISYEEDROGAEPKRVKNETKTYFKVQHNMAPTQKDFCDKAWAYFSDV 1860
Db 1801 PYSFYSSLIISYEEDROGAEPKRVKNETKTYFKVQHNMAPTQKDFCDKAWAYFSDV 1860
Qy 1861 DLEKDVHSLGLPLVCHTNTLNPAHGQVTVQEPALFTTIDETKSWYFTENMERNCRA 1920
Db 1861 DLEKDVHSLGLPLVCHTNTLNPAHGQVTVQEPALFTTIDETKSWYFTENMERNCRA 1920
Qy 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVMAQDQIRIRWYLLSGMGNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVMAQDQIRIRWYLLSGMGNENIHSIHFSGH 1980
Qy 1981 VFTVKKSEYKMAJYNYLPQGVFTEVEMLPKAGIWRVECLIGELHLAGNSTLFLVYSNKC 2040
Db 1981 VFTVKKSEYKMAJYNYLPQGVFTEVEMLPKAGIWRVECLIGELHLAGNSTLFLVYSNKC 2040
Qy 2041 QTPLGWASGHIRDFOITASQYQGWAPKIALRHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPLGWASGHIRDFOITASQYQGWAPKIALRHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Qy 2101 HGIKTQARQKFSYIYSOFIIMYSLDGKKWQTYRGNSGTGTLWFFGNVDSSGIKHNTFN 2160
Db 2101 HGIKTQARQKFSYIYSOFIIMYSLDGKKWQTYRGNSGTGTLWFFGNVDSSGIKHNTFN 2160
Qy 2161 PPIIARYIRLPHTHYSIRSTLRMLMGCDLNCSPMLGSKAISDAQTASSYFTNMEFA 2220
Db 2161 PPIIARYIRLPHTHYSIRSTLRMLMGCDLNCSPMLGSKAISDAQTASSYFTNMEFA 2220
Qy 2221 TWSPSKARLHLOGRSNAPQVNNPKWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAPQVNNPKWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKRVKVFQGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRM 2340

Db 2281 ISSSQDGHQWTLFFQNGKRVKVFQGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRM 2340
Qy 2341 EVLGCENQDLY 2351
Db 2341 EVLGCENQDLY 2351
RESULT 6
ABB99528
ID ABB99528 standard; protein; 2351 AA.
AC ABB99528;
XX
DT 03-MAR-2003 (first entry)
XX
DE Amino acid sequence of human Factor VIII.
XX
KW Factor VIII; non-viral vesicle vector; vesicular membrane; hepatitis B;
XX envelope protein; Factor IX; liver cell; gene therapy; haemophilia.
XX Homo sapiens.
OS
PN WO200286091-A2.
XX
PD 31-OCT-2002.
XX
PF 25-APR-2002; 2002WO-US013164.
XX
PR 25-APR-2001; 2001US-0286314P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chien KR, Hoshijima M;
XX
DR WPI; 2003-093125/08.
XX
DR N-PSDB; ABV77052.
XX
PT New non-viral vesicle vector comprises vesicular membrane with hepatitis
PT B envelope protein and nucleic acid expression construct comprising
PT complete factor VIII or IX coding sequence, useful for treating
PT hemophilia.
XX
PS Disclosure; Page 22-32; 34pp; English.
XX
CC The present sequence represents a human Factor VIII. Factor VIII
CC polynucleotides are used to construct non-viral vesicle vectors. These
CC vectors comprise a vesicular membrane with hepatitis B envelope protein
CC exposed on the vesicle surface, and a nucleic acid expression construct
CC comprising a complete factor VIII or factor IX coding sequence and a
CC promoter sequence functional in liver cells. The non-viral vesicle vector
CC is useful in gene therapy protocols for treating haemophilia
XX
SQ Sequence 2351 AA;
Query Match 100.0%; Score 12410; DB 6; Length 2351;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MOIELSTCFELCLLFCFSATRRYYLGAVELSWDVMQSDLGELPYDARFPPRPVKSFPFN 60
Db 1 MOIELSTCFELCLLFCFSATRRYYLGAVELSWDVMQSDLGELPYDARFPPRPVKSFPFN 60
Qy 61 TSVVYKTLFVEFTDHLFNIAKRPFPWMLLGPITQAEVYDTVTTLKMAASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDHLFNIAKRPFPWMLLGPITQAEVYDTVTTLKMAASHPVSLHAV 120
Qy 121 GVSYWASSEGAEYDDQTSQREKEDKVPFGSGHTTVVQVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSYWASSEGAEYDDQTSQREKEDKVPFGSGHTTVVQVLKENGPMASDPLCLTYSYLH 180
Qy 181 VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFTLLFAVFDGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFTLLFAVFDGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPMHTVNGYVARSUGLIGCHRSKYVWHVIGWGTTPVHVSIFLHGHTFLVRNH 300
DB 241 AASARAWPMHTVNGYVARSUGLIGCHRSKYVWHVIGWGTTPVHVSIFLHGHTFLVRNH 300
QY 301 ROASLEISPIITLTAQTLMDLQFLLSCHLSSHQDGMAYVYKVDSCPEPQLRMKNE 360
DB 301 ROASLEISPIITLTAQTLMDLQFLLSCHLSSHQDGMAYVYKVDSCPEPQLRMKNE 360
QY 361 EABYDDDLTDSMDVPRDDDNSSFTQIRSAVAKHFKTWHTWHTAAABEDWDYAPLVLA 420
DB 361 EABYDDDLTDSMDVPRDDDNSSFTQIRSAVAKHFKTWHTWHTAAABEDWDYAPLVLA 420
QY 421 PDRSVKSOYLNNGPQIRGRKYKVRFWAYTDETFKTREAIQHESGILGPLLYGVBGDTL 480
DB 421 PDRSVKSOYLNNGPQIRGRKYKVRFWAYTDETFKTREAIQHESGILGPLLYGVBGDTL 480
QY 481 LIIFXNQASRPNIYPHGTDVRLYSRRLPKGVKHLKDFPLPGBEIPKYKWTVTVEDGP 540
DB 481 LIIFXNQASRPNIYPHGTDVRLYSRRLPKGVKHLKDFPLPGBEIPKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSYFVNMEDLASGLIGPLLIYKESVDQORNOIMSDKRNVLFSVFE 600
DB 541 TKSDPRCLTRYSSYFVNMEDLASGLIGPLLIYKESVDQORNOIMSDKRNVLFSVFE 600
QY 601 NRSWLTENIQRFPLNPAGVQLEDPEFOASNMHSINGVVPDSLOLSVCLHEVAVWYILS 660
DB 601 NRSWLTENIQRFPLNPAGVQLEDPEFOASNMHSINGVVPDSLOLSVCLHEVAVWYILS 660
QY 661 IGAQTDFLSVFSGVTFKHVMYEDTLTLPFSGETVFMNMENPGLWILGCHNSPFRNG 720
DB 661 IGAQTDFLSVFSGVTFKHVMYEDTLTLPFSGETVFMNMENPGLWILGCHNSPFRNG 720
QY 721 YTMALIVSSCDKNTGVDYEDISAYLLSKNALEPRSFSONSHPSFROKQFNATTI 780
DB 721 YTMALIVSSCDKNTGVDYEDISAYLLSKNALEPRSFSONSHPSFROKQFNATTI 780
QY 781 PENDIEKTPWFAHRTPMKIQNVSSDLMILLRQSPFHGILSLDQAKYETTSDDPS 840
DB 781 PENDIEKTPWFAHRTPMKIQNVSSDLMILLRQSPFHGILSLDQAKYETTSDDPS 840
QY 841 PGALDNNLSLSEMTHPOLHSGDMVFTPEGLQLRLNEKIGTTAATLKKLDFKVSST 900
DB 841 PGALDNNLSLSEMTHPOLHSGDMVFTPEGLQLRLNEKIGTTAATLKKLDFKVSST 900
QY 901 SNKLSTIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLFGKSSPLTSEGGLPSLEE 960
DB 901 SNKLSTIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLFGKSSPLTSEGGLPSLEE 960
QY 961 NNDKSLLESGLMNSQSSWGKNVSSSTESGRLPKGAHGPALLTKDNALFKVYSISLLKTN 1020
DB 961 NNDKSLLESGLMNSQSSWGKNVSSSTESGRLPKGAHGPALLTKDNALFKVYSISLLKTN 1020
QY 1021 KYSNNSATNRKTHIDGPSLLIENSPPVWQNILESDETFPKVTPFLIHDRMLMDKNATLRL 1080
DB 1021 KYSNNSATNRKTHIDGPSLLIENSPPVWQNILESDETFPKVTPFLIHDRMLMDKNATLRL 1080
QY 1081 NMSNKTTSKMNEMVQCKEGEIPDPAQNPDMSEFFKMLFLPESARWIORTHGKNSLNG 1140
DB 1081 NMSNKTTSKMNEMVQCKEGEIPDPAQNPDMSEFFKMLFLPESARWIORTHGKNSLNG 1140
QY 1141 QGSPKQVSLGPEKSVGEGQNFISEKXNVVVGKEFTKDVGLKEMVFPSSRNFLTNLDN 1200
DB 1141 QGSPKQVSLGPEKSVGEGQNFISEKXNVVVGKEFTKDVGLKEMVFPSSRNFLTNLDN 1200
QY 1201 LHENNTNQBKKIOBEIEKKEKTHILOENVLPQHTVTGTGNFMKNLFLLSSTRQNVESYD 1260
DB 1201 LHENNTNQBKKIOBEIEKKEKTHILOENVLPQHTVTGTGNFMKNLFLLSSTRQNVESYD 1260
QY 1261 GAYAPVLQDFRSKNDSTNRKTHATFSSKKEEENLEGLNQTKQIVKVCATTRISPN 1320
DB 1261 GAYAPVLQDFRSKNDSTNRKTHATFSSKKEEENLEGLNQTKQIVKVCATTRISPN 1320

QY 1321 SQQNFVTRSKRALKQFRLPLEETELEKRIIVDDTSIQWSKNMKHLPSTLTQIDYNEKE 1380
DB 1321 SQQNFVTRSKRALKQFRLPLEETELEKRIIVDDTSIQWSKNMKHLPSTLTQIDYNEKE 1380
QY 1381 KGALTQSPLSCLTRSHSIPOANRSPPIAKVSSPPSIRPIYLTRVLPODNSSHLPAASY 1440
DB 1381 KGALTQSPLSCLTRSHSIPOANRSPPIAKVSSPPSIRPIYLTRVLPODNSSHLPAASY 1440
QY 1441 RKQSGVQESSHFIQGAKKNNLSAAILTLEMTGQREVSLGTSATNSVTYKVKVENTVLP 1500
DB 1441 RKQSGVQESSHFIQGAKKNNLSAAILTLEMTGQREVSLGTSATNSVTYKVKVENTVLP 1500
QY 1501 KPDLPTKSGKVELLPKVHIYQKDLFPTTSGSPGHLDLVEGSLLOGTEGAIKWNEANRP 1560
DB 1501 KPDLPTKSGKVELLPKVHIYQKDLFPTTSGSPGHLDLVEGSLLOGTEGAIKWNEANRP 1560
QY 1561 GKVPFLRVATESSAKTSPSKLADPLAWDNHYGTQIPKEWKSOEKSPEKTAKKXDTILSL 1620
DB 1561 GKVPFLRVATESSAKTSPSKLADPLAWDNHYGTQIPKEWKSOEKSPEKTAKKXDTILSL 1620
QY 1621 NACSNHAIANAINEGONKPEIEVTWAKQRTERLCSQNPVVKHQRREITRTTLOSQDEE 1680
DB 1621 NACSNHAIANAINEGONKPEIEVTWAKQRTERLCSQNPVVKHQRREITRTTLOSQDEE 1680
QY 1681 IDYDPTISVEMKEDFDIYDEDNQSPSPFOKKTSHYFIAAVERLWYGMSSSPHVLNR 1740
DB 1681 IDYDPTISVEMKEDFDIYDEDNQSPSPFOKKTSHYFIAAVERLWYGMSSSPHVLNR 1740
QY 1741 AQSGSVPOPKVWFOEFTDGSFTQPLYRGELNEHLGGLGPYIRAEVEDNIWVTFNQASR 1800
DB 1741 AQSGSVPOPKVWFOEFTDGSFTQPLYRGELNEHLGGLGPYIRAEVEDNIWVTFNQASR 1800
QY 1801 PYSYSSLIISYEEDORQGAEPKRNKFNKTYFWKQHHMAPTKDFDCKAWAYFSDV 1860
DB 1801 PYSYSSLIISYEEDORQGAEPKRNKFNKTYFWKQHHMAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGILGPLLVCHTNTLNPAGHQRVTVQEPALFPTTIDETKSYFYTENMERNCA 1920
DB 1861 DLEKDVHSLGILGPLLVCHTNTLNPAGHQRVTVQEPALFPTTIDETKSYFYTENMERNCA 1920
QY 1921 PCNIQMEDPTKENYRPHAINGYIMDTLPGLVMAQDQIRKAYLLSMGSENIHSTHSGH 1980
DB 1921 PCNIQMEDPTKENYRPHAINGYIMDTLPGLVMAQDQIRKAYLLSMGSENIHSTHSGH 1980
QY 1981 VFTVRKSEYKMALYNLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKSEYKMALYNLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPLGASGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKBPFFSWIKVDLLAPMII 2100
DB 2041 QTPLGASGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKBPFFSWIKVDLLAPMII 2100
QY 2101 HGKIQGARQKSSLYISQFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSGIGHNIFN 2160
DB 2101 HGKIQGARQKSSLYISQFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSGIGHNIFN 2160
QY 2161 PIIARIYIRLHPHTHYSIRSLTLEMLMGCDLNSCSPGLMGESKAIQITASSYFTNMPA 2220
DB 2161 PIIARIYIRLHPHTHYSIRSLTLEMLMGCDLNSCSPGLMGESKAIQITASSYFTNMPA 2220
QY 2221 TWSPEKARLHLOGRNARVQNNPKEMI QVDFQTKMKTGVTTQGVKSLTSMVVKFEL 2280
DB 2221 TWSPEKARLHLOGRNARVQNNPKEMI QVDFQTKMKTGVTTQGVKSLTSMVVKFEL 2280
QY 2281 ISSSQDGHOWTLFFONGKVKVFCGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRM 2340
DB 2281 ISSSQDGHOWTLFFONGKVKVFCGNQDSFTPVVNSLDPLLTRYLRIHPQSWHQAIALRM 2340
QY 2341 EYLGCEAQDLY 2351
DB 2341 EYLGCEAQDLY 2351

RESULT 7

AAW11416
 ID AAW11416 standard; protein; 2351 AA.
 XX
 AC AAW11416;
 DT 20-NOV-1997 (first entry)
 XX
 DE Active Factor VIII:C analogue A774X.
 KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;
 KW proteolytic cleavage.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key
 FT Peptide
 FT /note= "signal peptide"
 FT 20..2351
 FT /note= "mature Factor VIII:C"
 FT 20..1667
 FT /note= "heavy chain fragment"
 FT 760..1667
 FT /note= "B domain"
 FT 774
 FT Modified-site
 FT /label= Phe, Glu, Pro
 FT 1668..2350
 FT Region
 FT /note= "light chain fragment"
 XX
 DN W03703195-A1.
 XX
 PD 30-JAN-1997.
 XX
 PF 09-JUL-1996; 96WO-US011444.
 XX
 PR 11-JUL-1995; 95US-0001025P.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Hung DT, Cohen FE, Innis M;
 XX
 XX WPI; 1997-119050/11.
 XX
 DR Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 PT used in the treatment of haemophiliacs, by improvement of haemostasis.
 PT
 XX Claim 26; Page; 90pp; English.
 PS
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention.. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAW1357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;

Query Watch 99.9%; Score 12406; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSDGLGELPVDARFPFRVPKSPFFN 60
 DB 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSDGLGELPVDARFPFRVPKSPFFN 60
 QY 61 TSVVYKXTLFEFTDHLFNIAKPRPPWMLGPTIQAEVVDVTVITLKNASHVSLHAV 120
 DB 61 TSVVYKXTLFEFTDHLFNIAKPRPPWMLGPTIQAEVVDVTVITLKNASHVSLHAV 120
 QY 121 GVSYWKASGAEYDDQTSQREKDDKVPFGSSHTYVWQVLKENGPMASDPLCLTYSLSH 180
 DB 121 GVSYWKASGAEYDDQTSQREKDDKVPFGSSHTYVWQVLKENGPMASDPLCLTYSLSH 180
 QY 181 VDLVKDNLNSGLIGALLVCREGSLAKEKFTQLHKFILLFAVDEBKSHSETKNSLMQDRD 240
 DB 181 VDLVKDNLNSGLIGALLVCREGSLAKEKFTQLHKFILLFAVDEBKSHSETKNSLMQDRD 240
 QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRAKSVYWHVIGMGTTPVHSLFLECHTFLVRNH 300
 DB 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRAKSVYWHVIGMGTTPVHSLFLECHTFLVRNH 300
 QY 301 ROASLEISPIITFLTAQTLMLDGLQFLISCHTSSHOHDGMEAYKVVDSCPEEPQOLRMKNE 360
 DB 301 ROASLEISPIITFLTAQTLMLDGLQFLISCHTSSHOHDGMEAYKVVDSCPEEPQOLRMKNE 360
 QY 361 EAEYDDDLTDSEMDVVRPDDNSPSTQIISVAKKPKTWHTYIAABEEDWDZAPLVLA 420
 DB 361 EAEYDDDLTDSEMDVVRPDDNSPSTQIISVAKKPKTWHTYIAABEEDWDZAPLVLA 420
 QY 421 PDDRSYKSOYLNGEFORIGRYKVRNWAYTDEFKTRKREAIQHESGLIGLLEGVGDTL 480
 DB 421 PDDRSYKSOYLNGEFORIGRYKVRNWAYTDEFKTRKREAIQHESGLIGLLEGVGDTL 480
 QY 481 LLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEGDP 540
 DB 481 LLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEGDP 540
 QY 541 TKSDPRCLTRYSSPVNMRDLASGLIGPLIICYKESVDQGNQIMSKRNVILFSVDE 600
 DB 541 TKSDPRCLTRYSSPVNMRDLASGLIGPLIICYKESVDQGNQIMSKRNVILFSVDE 600
 QY 601 NRSWYLTENIQRFENPAGVQLEDPEFQASNMHSINGVVFDSLSQSVCLHEVAWYILS 660
 DB 601 NRSWYLTENIQRFENPAGVQLEDPEFQASNMHSINGVVFDSLSQSVCLHEVAWYILS 660
 QY 661 IGAQTDFLSVFPGYTFKHQWVYEDTLTLPFPGSETVFMSEMENPGLWILGCHNSDFNRG 720
 DB 661 IGAQTDFLSVFPGYTFKHQWVYEDTLTLPFPGSETVFMSEMENPGLWILGCHNSDFNRG 720
 QY 721 MTALKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHSHTROKQFNATTI 780
 DB 721 MTALKVSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHSHTROKQFNATTI 780
 QY 781 PENDIEKTDPMFAHETPMKIQNVSSDLLMLLROSPTPHGLSLSDLOEAKYETESDDPS 840
 DB 781 PENDIEKTDPMFAHETPMKIQNVSSDLLMLLROSPTPHGLSLSDLOEAKYETESDDPS 840
 QY 841 PGADSNNSLSEMTFRPQLHSHSGDMVTTPESGQLRLNEKLGTTAATLKKLDFKVSST 900
 DB 841 PGADSNNSLSEMTFRPQLHSHSGDMVTTPESGQLRLNEKLGTTAATLKKLDFKVSST 900
 QY 901 SNNLSTTIPSDNLAAGTNTSLSGPPSMVHYDSOLDITLFGKKSPLTESGPPSLSEE 960
 DB 901 SNNLSTTIPSDNLAAGTNTSLSGPPSMVHYDSOLDITLFGKKSPLTESGPPSLSEE 960
 QY 961 NNDSKLLSGLMNSQSSWGKNVSTESGRLFKGRARHGAPALLTKDNALFKVSIULLKTN 1020
 DB 961 NNDSKLLSGLMNSQSSWGKNVSTESGRLFKGRARHGAPALLTKDNALFKVSIULLKTN 1020

QY 1021 KTSNNSATNRKTHIDGPFLLIENSFVQWNILESDETEKKTPTLHDMXMDKXATLRL 1080
DB 1021 KTSNNSATNRKTHIDGPFLLIENSFVQWNILESDETEKKTPTLHDMXMDKXATLRL 1080
QY 1081 NEMSNKTTSSKNMVMQOKKEGPIPPDAQNPDMSFFKMLFJPESARWQRTHGKXSLNSG 1140
DB 1081 NEMSNKTTSSKNMVMQOKKEGPIPPDAQNPDMSFFKMLFJPESARWQRTHGKXSLNSG 1140
QY 1141 QGSPKQVLSLGPERSVEQONFLSKNKVAVCKGFTKDVGKEMVFPSSRNLFITNLDN 1200
DB 1141 QGSPKQVLSLGPERSVEQONFLSKNKVAVCKGFTKDVGKEMVFPSSRNLFITNLDN 1200
QY 1201 LHENNTHQEKKIQSEIEKKEKTELIGENVVLPOIHTVGTGKFMKMLFLLSTRQNVGSSVD 1260
DB 1201 LHENNTHQEKKIQSEIEKKEKTELIGENVVLPOIHTVGTGKFMKMLFLLSTRQNVGSSVD 1260
QY 1261 GAYAPVLQDFRSLNDSTNRKXHTAHFSKKGSEENLEGLNQTOKIOIVKIACTTRISNT 1320
DB 1261 GAYAPVLQDFRSLNDSTNRKXHTAHFSKKGSEENLEGLNQTOKIOIVKIACTTRISNT 1320
QY 1321 SOONFVQSSKALKOFPLPESLETELEKELIYDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SOONFVQSSKALKOFPLPESLETELEKELIYDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGALTQSPSLDCITRSHSIFQANRSPPLPIAKVSSPFSIRPIYLTAVLFDQNSSHLPAASY 1440
DB 1381 KGALTQSPSLDCITRSHSIFQANRSPPLPIAKVSSPFSIRPIYLTAVLFDQNSSHLPAASY 1440
QY 1441 RKDQSGVQBSHFLOKAKKNNLSLAILTLEMTGDQREVSGLSATNSVYKKVENTVLP 1500
DB 1441 RKDQSGVQBSHFLOKAKKNNLSLAILTLEMTGDQREVSGLSATNSVYKKVENTVLP 1500
QY 1501 KEDLPKTSKVELLPKVHVIYQDLPFTETSNQSPGHLQVEGSLQGTGCAIKWNEANRP 1560
DB 1501 KEDLPKTSKVELLPKVHVIYQDLPFTETSNQSPGHLQVEGSLQGTGCAIKWNEANRP 1560
QY 1561 GKVPFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQEKSPKTAFAKGTILSL 1620
DB 1561 GKVPFLRVATSSAKTPSKLLDPLANDNHYGTQIPKEWKSQEKSPKTAFAKGTILSL 1620
QY 1621 NACESNHAIAANEGQNKPEIEVTWAKQGTERTCLSQNPVLRKHOREITRTTLQSDQEE 1680
DB 1621 NACESNHAIAANEGQNKPEIEVTWAKQGTERTCLSQNPVLRKHOREITRTTLQSDQEE 1680
QY 1681 IDYDDTISVEMKEDDIDYDDENQSPGFQKTRHYFIAAVERLDYGVSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKEDDIDYDDENQSPGFQKTRHYFIAAVERLDYGVSSSPHVLNR 1740
QY 1741 AQSGSVQPKKVVQFQFTDGSFTQPLRYGSELNHLGLGYTIRAEVEDNIMVTRNQASR 1800
DB 1741 AQSGSVQPKKVVQFQFTDGSFTQPLRYGSELNHLGLGYTIRAEVEDNIMVTRNQASR 1800
QY 1801 PYSFYSSLISYEDQDQGAEPKRNKVPNETKYFWKVQHMAPTDEPDCKAWAYFSDV 1860
DB 1801 PYSFYSSLISYEDQDQGAEPKRNKVPNETKYFWKVQHMAPTDEPDCKAWAYFSDV 1860
QY 1861 DLBKDVHSGLIGLLVCHTNTLNPAGHQVTVQEFALFFTFDETKSWYFTENMERNCR 1920
DB 1861 DLBKDVHSGLIGLLVCHTNTLNPAGHQVTVQEFALFFTFDETKSWYFTENMERNCR 1920
QY 1921 PCNIQMEDPTFKENYRPHANGIYIMDTPLGLVNAQDQRIWYLLSGNSNENIHSIHFSGH 1980
DB 1921 PCNIQMEDPTFKENYRPHANGIYIMDTPLGLVNAQDQRIWYLLSGNSNENIHSIHFSGH 1980
QY 1981 VFTVRKEEYKALNYLYPGVFEVEMLPKAGIWRVECLIGHLAGMSTFLVYSNKC 2040
DB 1981 VFTVRKEEYKALNYLYPGVFEVEMLPKAGIWRVECLIGHLAGMSTFLVYSNKC 2040
QY 2041 QTPLGASGHIRDFOITASGYQOWAPKAPLHVEGGINAWSTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLGASGHIRDFOITASGYQOWAPKAPLHVEGGINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQAROKFSSLYISQFIIMTSLDCKKWQTYRGNSGTGLMVFFGNVDSGIGKINIFN 2160

DB 2101 HGIKTQAROKFSSLYISQFIIMTSLDCKKWQTYRGNSGTGLMVFFGNVDSGIGKINIFN 2160
QY 2161 PPIIARYIRLPHTHYSIRSTRMLMGLDNLSCSMPLGWSKASDAQITASSYFTNMEA 2220
DB 2161 PPIIARYIRLPHTHYSIRSTRMLMGLDNLSCSMPLGWSKASDAQITASSYFTNMEA 2220
QY 2221 TWSPSKARLHLQGSNARWPQNNPKWLQVDFOKTMKVTVGTQGVKSLTSMYVKEFL 2280
DB 2221 TWSPSKARLHLQGSNARWPQNNPKWLQVDFOKTMKVTVGTQGVKSLTSMYVKEFL 2280
QY 2281 ISSODGCHQWTLFPONGKVKVQGNQGSFTPVVNSLDPPLLTRIRIHPOQSWHQAIALRM 2340
DB 2281 ISSODGCHQWTLFPONGKVKVQGNQGSFTPVVNSLDPPLLTRIRIHPOQSWHQAIALRM 2340
QY 2341 EVLGCBAODLY 2351
DB 2341 EVLGCBAODLY 2351
RESULT 8
AAW11343
ID AAW11343 standard; protein; 2351 AA.
XX AC AAW11343;
XX DT 17-NOV-1997 (first entry)
XX DE Active Factor VIII:C analogue S224X.
XX KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
XX KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
XX KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
XX XX proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
XX FH Key
XX FH Location/Qualifiers
FT Peptide 1..19
FT /note= "signal peptide"
FT Pro-sein 20..2351
FT /note= "mature Factor VIII:C"
FT Region 20..1667
FT /note= "heavy chain fragment"
FT Modified-site 243
FT /label= Phe, Glu, Pro
FT Domain 760..1667
FT /note= "B domain"
FT Region 1668..2350
FT /note= "light chain fragment"
XX WC9703195-A1.
XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR) CHIRON CORP.
XX Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophilias, by improvement of haemostasis.
XX Claim 10; Page; 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor

CC VIII:C coding sequence (see R451357) using mutagenic primers. The
 CC analogues comprise a native factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;

Query Match 99.9%; Score 2406; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOELSTCFCLARFCESATRYVYLGAVELSDYMQSDLGELPVDARFPERVPSRPPN 60
 DB J MOELSTCFCLARFCESATRYVYLGAVELSDYMQSDLGELPVDARFPERVPSRPPN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKRPWMLGPGTQAQVYTVITLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLFNIAKRPWMLGPGTQAQVYTVITLKNMASHPVSLHAV 120

QY 121 GVSWKASEGAEYDQTSQRKEDKVPFGSGHTYVWQVLKENGPMASDPLCLTVSYLSH 180
 DB 121 GVSWKASEGAEYDQTSQRKEDKVPFGSGHTYVWQVLKENGPMASDPLCLTVSYLSH 180

QY 181 VDLKNSGLTGALLVCRSGSLAKEKQTQLHKFILLFAVDEGKSHSETKNSLMQDRD 240
 DB 181 VDLKNSGLTGALLVCRSGSLAKEKQTQLHKFILLFAVDEGKSHSETKNSLMQDRD 240

QY 241 AASARAFKMHITVNGVYVNRSLGLICGHRKSYVHVIWGMITPEVHSIFLEGHTFLVRNH 300
 DB 241 AASARAFKMHITVNGVYVNRSLGLICGHRKSYVHVIWGMITPEVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPIITFLTAQTLMDLGOPLLSCHISSHODGMEAYVKVDSCPEBPOLMKNE 360
 DB 301 ROASLEISPIITFLTAQTLMDLGOPLLSCHISSHODGMEAYVKVDSCPEBPOLMKNE 360

QY 361 EAEDYDDDLTDSMDVRFDDNPSFTQIRSAVKKPKTWVHYTAAREEDWDYAPLVIA 420
 DB 361 EAEDYDDDLTDSMDVRFDDNPSFTQIRSAVKKPKTWVHYTAAREEDWDYAPLVIA 420

QY 421 PDRSYKQYLLNGPQRIKRYKVFWAYTDETEKTRAIQHESGIILGPILLYGEGDTL 480
 DB 421 PDRSYKQYLLNGPQRIKRYKVFWAYTDETEKTRAIQHESGIILGPILLYGEGDTL 480

QY 481 LIIFKQASRPNIYPHGITDVRPLYSRRLPKGVHLDKDFPLPGELIPKYKWTIVVEDGP 540
 DB 481 LIIFKQASRPNIYPHGITDVRPLYSRRLPKGVHLDKDFPLPGELIPKYKWTIVVEDGP 540

QY 541 TKSDPRCLTRYYSFVNMERDLASGLIGPLTICYKESVDORGNQIMSKRNVLFSVDE 600
 DB 541 TKSDPRCLTRYYSFVNMERDLASGLIGPLTICYKESVDORGNQIMSKRNVLFSVDE 600

QY 601 NRSWLTENIQRFLEPNAGVQLEDPEFOASNMHSINGVYVPSLQSLVCLHEVAYWYLS 660
 DB 601 NRSWLTENIQRFLEPNAGVQLEDPEFOASNMHSINGVYVPSLQSLVCLHEVAYWYLS 660

QY 661 ICAQTDPLSVFSGVTFKHWYEDTLTLPFSGETVFMNMPNGLWILGCHNSDFNRG 720
 DB 661 ICAQTDPLSVFSGVTFKHWYEDTLTLPFSGETVFMNMPNGLWILGCHNSDFNRG 720

QY 721 MTALLKVSSCKNTGDIYEDSYEDI SAYLLSKNAIEPRSFQNSRHPSTKQKFNATTI 780
 DB 721 MTALLKVSSCKNTGDIYEDSYEDI SAYLLSKNAIEPRSFQNSRHPSTKQKFNATTI 780

QY 781 PENDIEXTDWFPAHRTMPMKIQNVSSDILLMLLRQSPTRHGLSLSDIOEAKYETFSDDPS 840
 DB 781 PENDIEXTDWFPAHRTMPMKIQNVSSDILLMLLRQSPTRHGLSLSDIOEAKYETFSDDPS 840

QY 841 PGALDSNNSISEMTHFPQPLHSGDMVFTBESGLOELNEKLGTTAATELKKLDFKVSST 900
 DB 841 PGALDSNNSISEMTHFPQPLHSGDMVFTBESGLOELNEKLGTTAATELKKLDFKVSST 900

QY 901 SNNLITIPSDNLAAAGTNTSSLGPPMPVHYDSQDITLTFGKSSPLTSSGGPLSISEE 960
 DB 901 SNNLITIPSDNLAAAGTNTSSLGPPMPVHYDSQDITLTFGKSSPLTSSGGPLSISEE 960

QY 961 NNDKLLSGLMNSQESSWGKVSSTSGRLFKKRAHGAPALLTKDNALFKVSI SLKTN 1020
 DB 961 NNDKLLSGLMNSQESSWGKVSSTSGRLFKKRAHGAPALLTKDNALFKVSI SLKTN 1020

QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWOKILBDSDFKVTPLIHDRMLMDKNATLRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWOKILBDSDFKVTPLIHDRMLMDKNATLRL 1080

QY 1081 NMSNKTTSKQMEMVQKKEGPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLSNG 1140
 DB 1081 NMSNKTTSKQMEMVQKKEGPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLSNG 1140

QY 1141 QGPSFKQIVSLGPKSVGEGQNFISEKNVVGKGEFTQVGLKEMVFPSSNLFITNLDN 1200
 DB 1141 QGPSFKQIVSLGPKSVGEGQNFISEKNVVGKGEFTQVGLKEMVFPSSNLFITNLDN 1200

QY 1201 LHENNTNHOEKKIOEIEKKETLLOENVVLPOIHTVTCTKXFMKNLFLSTQNVEGSD 1260
 DB 1201 LHENNTNHOEKKIOEIEKKETLLOENVVLPOIHTVTCTKXFMKNLFLSTQNVEGSD 1260

QY 1261 GAYAPVLQDFPSLNDSTNRKTHAHFSKKEENLEGLNQTKQIVKAYACTRISNT 1320
 DB 1261 GAYAPVLQDFPSLNDSTNRKTHAHFSKKEENLEGLNQTKQIVKAYACTRISNT 1320

QY 1321 SQQNVFQTSKRALKQPLPLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQDYNKE 1380
 DB 1321 SQQNVFQTSKRALKQPLPLEETELEKRIIVDDTSTQSKNMKHLTPSTLTQDYNKE 1380

QY 1381 KGATQSPCLSRSHSIPQANSPPIAKVSPSPRIPIYLTVLFDQNSHLPAAASY 1440
 DB 1381 KGATQSPCLSRSHSIPQANSPPIAKVSPSPRIPIYLTVLFDQNSHLPAAASY 1440

QY 1441 RKDQSGVQESSHFLQGAKKNNLSLAITLMTGQREVGSGLGTSATNSVTYKVENTVLP 1500
 DB 1441 RKDQSGVQESSHFLQGAKKNNLSLAITLMTGQREVGSGLGTSATNSVTYKVENTVLP 1500

QY 1501 KPDLPKTSKVELLPKHIIYOKDLPPTSETSGSGHLDLVEGSLLOGTGAIKNEANRP 1560
 DB 1501 KPDLPKTSKVELLPKHIIYOKDLPPTSETSGSGHLDLVEGSLLOGTGAIKNEANRP 1560

QY 1561 GKVPFLRVATESSAKTSPKLLDPLAWDNHYGTQIPKEWKSQEKSPKTAFFKXDTILSL 1620
 DB 1561 GKVPFLRVATESSAKTSPKLLDPLAWDNHYGTQIPKEWKSQEKSPKTAFFKXDTILSL 1620

QY 1621 NACSNHAAIAINSGQNKPEIEVTWAKQGRTERLCSQNPVVLKRHOREITRTTLOSQDEE 1680
 DB 1621 NACSNHAAIAINSGQNKPEIEVTWAKQGRTERLCSQNPVVLKRHOREITRTTLOSQDEE 1680

QY 1681 IDYDDTISEVKKGBDFDIYDEDENQSPRSFOKTRHYPIAAVERLWDYGMSSSPHVLNR 1740
 DB 1681 IDYDDTISEVKKGBDFDIYDEDENQSPRSFOKTRHYPIAAVERLWDYGMSSSPHVLNR 1740

QY 1741 AQSSVPOFKVAVQEFIDGSGFTQPLRYGELNEHLGLLGPVIRAEVDENIMVTFRNOASR 1800
 DB 1741 AQSSVPOFKVAVQEFIDGSGFTQPLRYGELNEHLGLLGPVIRAEVDENIMVTFRNOASR 1800

QY 1801 PYSYSSLIISYEEDQORQGAEPKQNFVKPNETKTIVFWKVOHNAPTKDEFFCKAWAYFSDV 1860

```
Db 1801 PYSFYSSLLSVEEDRQRAEPKPKVKNKYTKYQVHNAVKDFDCKNAWYFSDV 1860
QY 1861 DLEKDVHSLGLPLLVCHTNTLNPAHQVQTVQEPALFFTFIDETKSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLGLPLLVCHTNTLNPAHQVQTVQEPALFFTFIDETKSWYFTENMERNCR 1920
QY 1921 PCNIOMEDPTKENYRFAINGYINDTLPLGVMAODQRIWVLLSMGSGNENIHSHPGSH 1980
Db 1921 PCNIOMEDPTKENYRFAINGYINDTLPLGVMAODQRIWVLLSMGSGNENIHSHPGSH 1980
QY 1981 VFTVKKKEEYKMALNLYPGVFTVEMLPFSKAGIWRVECLIGEHLHAGMSILFVYSNKC 2040
Db 1981 VFTVKKKEEYKMALNLYPGVFTVEMLPFSKAGIWRVECLIGEHLHAGMSILFVYSNKC 2040
QY 2041 QTPLQMASCHIRDFQITASQYQWAPKARLHYSGSINAMSTKSPFSWIKVDLLAPMII 2100
Db 2041 QTPLQMASCHIRDFQITASQYQWAPKARLHYSGSINAMSTKSPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQGAQKFSLLYSIQFIIMYSIDGKWOTYGNSTGTLWVFPNGVDSGIGKHNFN 2160
Db 2101 HGIKTQGAQKFSLLYSIQFIIMYSIDGKWOTYGNSTGTLWVFPNGVDSGIGKHNFN 2160
QY 2161 PPIIARYIRLHPHTHSYISIRSLRMELMGCDLNSCSPLGMESKAI SDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPHTHSYISIRSLRMELMGCDLNSCSPLGMESKAI SDAQITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLOGRGNARPOVNNPKWLVQDFQTKMKVGTGTTQGVKSLTSMYVKZPL 2280
Db 2221 TWSPSKARLHLOGRGNARPOVNNPKWLVQDFQTKMKVGTGTTQGVKSLTSMYVKZPL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRHPQSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 9
AAW11435
ID AAW11435 standard; protein; 2351 AA.
XX
AC AAW11435;
DF 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue S1311X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "signal peptide"
FT Region /note= "mature Factor VIII:C"
FT Domain /note= "heavy chain fragment"
FT Modified-site 760..1667
FT Region /note= "B domain"
FT Region 1330
FT Region 1668..2350
FT Region /note= "light chain fragment"
XX
FN MO9703195-A1.
```

```
XX 30-JAN-1997.
PD
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR ) CHIRON CORP.
XX Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophilias, by improvement of haemostasis.
XX Claim 30; Page; 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophilias, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;
XX
Query Match 99.9%; Score 12406; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOELSTCFELCLRFCSATRRYYLGAVELSWDYNQSDLGELPVDARPPVPKSPFN 60
Db 1 MOELSTCFELCLRFCSATRRYYLGAVELSWDYNQSDLGELPVDARPPVPKSPFN 60
QY 61 TSVVYKTLFVEFTDHLNIAKPRPPNMGLLGPTTQAEVYDVTVTILNMASHPVSLHAV 120
Db 61 TSVVYKTLFVEFTDHLNIAKPRPPNMGLLGPTTQAEVYDVTVTILNMASHPVSLHAV 120
QY 121 GVSVMKASEGAEVDDQTSQREKDDKVPFGSGHTYVWVLKXENGPMASDPLCLTYSYLSH 180
Db 121 GVSVMKASEGAEVDDQTSQREKDDKVPFGSGHTYVWVLKXENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDFEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKOHTVNGYVNRSLPGLIGCHRKSVYVWVIGMGTTPPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKOHTVNGYVNRSLPGLIGCHRKSVYVWVIGMGTTPPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELTAQTLLMDLQGLLSCHSHOHGDMGAEVYKVDSCPEEPQLRMKNE 360
Db 301 ROASLEISPIITELTAQTLLMDLQGLLSCHSHOHGDMGAEVYKVDSCPEEPQLRMKNE 360
QY 361 EAEDYDDDLTDGEMDVRFDDDDNSPSFQIRSVAKKHPTWVHYTAAABEEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDGEMDVRFDDDDNSPSFQIRSVAKKHPTWVHYTAAABEEDWDYAPLVLA 420
```

QY 421 PDRS YKSOYLNNQPORIGSKYKVRMAWYDTEKTRERAIQHSGLGLPGLYGEVGDTL 480
DB 421 PDRS YKSOYLNNQPORIGSKYKVRMAWYDTEKTRERAIQHSGLGLPGLYGEVGDTL 480
QY 481 LIIFKQASRYNIYPHGIDTVPLRRPLFGVKKHLKDRPPLFGEIFKYKWTVTVEDGP 540
DB 481 LIIFKQASRYNIYPHGIDTVPLRRPLFGVKKHLKDRPPLFGEIFKYKWTVTVEDGP 540
QY 541 TKGDPRLCTIYYSFVNMERDLASGLIGPLLCYKESVDORGNQIMSDKRNVLFSVFDE 600
DB 541 TKGDPRLCTIYYSFVNMERDLASGLIGPLLCYKESVDORGNQIMSDKRNVLFSVFDE 600
QY 601 NRSWLTENIQRFPLNPAGVLEDBFQASNIMESINGYVDFSQLSVCLHEVAYWILS 660
DB 601 NRSWLTENIQRFPLNPAGVLEDBFQASNIMESINGYVDFSQLSVCLHEVAYWILS 660
QY 661 IGAOTDLSVFFSGYTFKHKKVYEDTLTLPFSGETVFMSENGLMILGCHNSDFNRG 720
DB 661 IGAOTDLSVFFSGYTFKHKKVYEDTLTLPFSGETVFMSENGLMILGCHNSDFNRG 720
QY 721 MTALLKVSCKDNTGDIYEDSYEDISAYLLSKNAIAPRSFSONSRHPSRQKQFNATTI 780
DB 721 MTALLKVSCKDNTGDIYEDSYEDISAYLLSKNAIAPRSFSONSRHPSRQKQFNATTI 780
QY 781 PENDIEKTDPAFAHRTMPKIQNVSSDMLLRQSPHPGLSLSDLOEAKYETFSDDPS 840
DB 781 PENDIEKTDPAFAHRTMPKIQNVSSDMLLRQSPHPGLSLSDLOEAKYETFSDDPS 840
QY 841 PGAI DSNNSLSEMTFRPQLAHSGDMVTFPSGQLRLNEKLGTTAATLKKLDFKVSST 900
DB 841 PGAI DSNNSLSEMTFRPQLAHSGDMVTFPSGQLRLNEKLGTTAATLKKLDFKVSST 900
QY 901 SNNLISITPSNLAAGTNTSGLPPSPNVHYDSOLDTTLFGKSSPLTESGGPLSSEE 960
DB 901 SNNLISITPSNLAAGTNTSGLPPSPNVHYDSOLDTTLFGKSSPLTESGGPLSSEE 960
QY 961 NNDKSLBESGLMNSQSSWGNVSTESGRFLFKGRAHGPALLTKDNALFKVSISSLKTN 1020
DB 961 NNDKSLBESGLMNSQSSWGNVSTESGRFLFKGRAHGPALLTKDNALFKVSISSLKTN 1020
QY 1021 KTSNNSATNRKTHIDCPSELLIENSFVWONILBSTDTEPKKVTPLIHDRMLMDXNATLRL 1080
DB 1021 KTSNNSATNRKTHIDCPSELLIENSFVWONILBSTDTEPKKVTPLIHDRMLMDXNATLRL 1080
QY 1081 NNMKNKTTSSKNMEXVQKKEGPIPDQAQNPMSFFKMLFIPESARWORTHGKNSLNSG 1140
DB 1081 NNMKNKTTSSKNMEXVQKKEGPIPDQAQNPMSFFKMLFIPESARWORTHGKNSLNSG 1140
QY 1141 QCPSPKQVSLGPEKSVESQNFLEKKNVVGKGFETKDVGLKEMVFPSSRNLFILTNLDN 1200
DB 1141 QCPSPKQVSLGPEKSVESQNFLEKKNVVGKGFETKDVGLKEMVFPSSRNLFILTNLDN 1200
QY 1201 LHENNTNHOEKI QBEIEKKETLIIQENVVLPOIHVTVGTGKNPKNLFILSTRQNVESYD 1260
DB 1201 LHENNTNHOEKI QBEIEKKETLIIQENVVLPOIHVTVGTGKNPKNLFILSTRQNVESYD 1260
QY 1261 GAYAVLQDFPSLNDSTNRKKGTAHFSKKGSEENLEGLGNQTKQIVKACTTIRISNPT 1320
DB 1261 GAYAVLQDFPSLNDSTNRKKGTAHFSKKGSEENLEGLGNQTKQIVKACTTIRISNPT 1320
QY 1321 SQONFVQTSKRALQFRLPLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQONFVQTSKRALQFRLPLEETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAIQTOSPLSDCLTRSHSIPQANRSLPIAKVSSPFSIRPIYLTTRVLFPQDNSHLPAASY 1440
DB 1381 KGAIQTOSPLSDCLTRSHSIPQANRSLPIAKVSSPFSIRPIYLTTRVLFPQDNSHLPAASY 1440
QY 1441 RKKDSGVQSSSHFLOCAKKNLSLALTLEMTGDQREVLGSLGTSATNSVTYKKVENTVLP 1500
DB 1441 RKKDSGVQSSSHFLOCAKKNLSLALTLEMTGDQREVLGSLGTSATNSVTYKKVENTVLP 1500
QY 1501 KPDLPTSGKVELLPKVHIYQKDLFPETSTNSGPHGLDLVEGSLLOCTEGAIAKWEANRP 1560

DB 1501 KPDLPTSGKVELLPKVHIYQKDLFPETSTNSGPHGLDLVEGSLLOCTEGAIAKWEANRP 1560
QY 1561 GKVPFLAVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQBSPEKTAFAKKDDTLISL 1620
DB 1561 GKVPFLAVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQBSPEKTAFAKKDDTLISL 1620
QY 1621 NACESNHAIAINNEGQNKPEIEVTWAKQGRTERLCSQNPVLAHRHQREIIRTTILSQDOEE 1680
DB 1621 NACESNHAIAINNEGQNKPEIEVTWAKQGRTERLCSQNPVLAHRHQREIIRTTILSQDOEE 1680
QY 1681 IDYDDTISVMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGHSSSPHVLRNR 1740
DB 1681 IDYDDTISVMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGHSSSPHVLRNR 1740
QY 1741 AQSGSVPOFKKVVQFBTQDGSFTQPLVRGELNHLGLGPYIRAEVEDNIMVFRNQASR 1800
DB 1741 AQSGSVPOFKKVVQFBTQDGSFTQPLVRGELNHLGLGPYIRAEVEDNIMVFRNQASR 1800
QY 1801 PYSFYSSLSIYEBEDQROGAEPKRFVKNPNETKITVFWKVOHHMPTKDEDFCKAWAYFSDV 1860
DB 1801 PYSFYSSLSIYEBEDQROGAEPKRFVKNPNETKITVFWKVOHHMPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGLIGPLLVCHTNTLPAHGRQVTVQEPALFFTFIDBTCKWYFTENNERCRA 1920
DB 1861 DLEKDVHSGLIGPLLVCHTNTLPAHGRQVTVQEPALFFTFIDBTCKWYFTENNERCRA 1920
QY 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQDRIWYLLSMGSNENIHSIHPSGH 1980
DB 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQDRIWYLLSMGSNENIHSIHPSGH 1980
QY 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPGLMASGHIRDPQITASGOYGQWAPKRLARLHYSGSINAWSTKEPSPWIKVDLLAPMII 2100
DB 2041 QTPGLMASGHIRDPQITASGOYGQWAPKRLARLHYSGSINAWSTKEPSPWIKVDLLAPMII 2100
QY 2101 HGIKTQAROKFESSIYISQFIIMYSLDGKKQTVRGNSTGTLMVFFGNVDSSGIKHNIEN 2160
DB 2101 HGIKTQAROKFESSIYISQFIIMYSLDGKKQTVRGNSTGTLMVFFGNVDSSGIKHNIEN 2160
QY 2161 PPIIARVIRLPTHYISIRSTLRMELMGCDLNSCMPLGMSKASISDAQITASSYFTNMFA 2220
DB 2161 PPIIARVIRLPTHYISIRSTLRMELMGCDLNSCMPLGMSKASISDAQITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLQGRSNARWPQVNNPKEWLOVDFOKTKMVTGVTTCQVKSLLTSMVVKFPL 2280
DB 2221 TWSPSKARLHLQGRSNARWPQVNNPKEWLOVDFOKTKMVTGVTTCQVKSLLTSMVVKFPL 2280
QY 2281 ISSSDQGHQWTLFFQNGKVKVFGQNDQSFTPVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
DB 2281 ISSSDQGHQWTLFFQNGKVKVFGQNDQSFTPVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
QY 2341 EYLGCEAQDLY 2351
DB 2341 EYLGCEAQDLY 2351
RESULT 10
AAW11445
ID AAW11445 standard; protein; 2351 AA.
XX
AC AAW11445;
XX
DT 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L1643X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;

KW	proteolytic cleavage.	QY	121	GVSVKASGEAEYDDDTSQREKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSH	180
XX		Db	121	GVSVKASGEAEYDDDTSQREKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSH	180
OS	Homo sapiens.	QY	181	VDLVKDLNSGLIGALLVCREGSLAKETQTLHKFILLFAVFDGKSWHSTKNSLMODRD	240
OS	Synthetic.	Db	181	VDLVKDLNSGLIGALLVCREGSLAKETQTLHKFILLFAVFDGKSWHSTKNSLMODRD	240
XX	Key	QY	241	AASARAWPKKHTVNGYVNRSLPGLIGCHRSVYVHWI GMGTTPVHIFLEGHTFLVRNH	300
FT	Peptide			/note= "signal peptide"	
FT	Protein	Db	241	AASARAWPKKHTVNGYVNRSLPGLIGCHRSVYVHWI GMGTTPVHIFLEGHTFLVRNH	300
FT	Region	QY	301	QOASLEISPIITFLPAQTLLMDLGOFLSCHSISSHQHDMGMEAYVKVDCSPPQPMKNE	360
FT	Domain	Db	301	QOASLEISPIITFLPAQTLLMDLGOFLSCHSISSHQHDMGMEAYVKVDCSPPQPMKNE	360
FT	Modified-site	QY	361	EAEYDDDLTDSMDVYRFDDNDSPSFIQIRSAVKHPKTWVHIAAEEDDWDVAPLVLA	420
FT	Region	Db	361	EAEYDDDLTDSMDVYRFDDNDSPSFIQIRSAVKHPKTWVHIAAEEDDWDVAPLVLA	420
XX	WO9703195-Al.	QY	421	PDRSYKSOYLNGPQIRIGKYKVRPMAVTDFTKTREAIQHESGILGPLLYGEVGDTL	480
XX	30-JAN-1997.	Db	421	PDRSYKSOYLNGPQIRIGKYKVRPMAVTDFTKTREAIQHESGILGPLLYGEVGDTL	480
XX	09-JUL-1996; 95WO-US011444.	QY	481	LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFPKYKWTVVEDGP	540
XX	11-JUL-1995; 95US-0001025P.	Db	481	LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFPKYKWTVVEDGP	540
XX	(CHIR) CHIRON CORP.	QY	541	TKSDPRCLTRYSSFNMRDLASGLIGPLI CYKESVDORGNQIMSDKRVILFSVFDE	600
XX	Hung DT, Cohen FE, Innis M;	Db	541	TKSDPRCLTRYSSFNMRDLASGLIGPLI CYKESVDORGNQIMSDKRVILFSVFDE	600
XX	WPI; 1997-119050/11.	QY	601	NRSWLTENIORFLPNPAGVQLEDPQASNMHSINGYVDSLSQLSVCLHEVAYWILS	660
XX	Factor VIII:C analog modified adjacent to a non-activating Arg residue -	Db	601	NRSWLTENIORFLPNPAGVQLEDPQASNMHSINGYVDSLSQLSVCLHEVAYWILS	660
XX	used in the treatment of haemophiliacs, by improvement of haemostasis.	QY	661	IGAQTDFLSVFPFGYTFKHKMYVEDTLTLPFPGSETVPMENGLWILGCHNSDFNRG	720
XX	Claim 32; Page: 90pp; English.	Db	661	IGAQTDFLSVFPFGYTFKHKMYVEDTLTLPFPGSETVPMENGLWILGCHNSDFNRG	720
XX	AAW11330-W11472 represent active Factor VIII:C analogues of the	QY	721	MTALKVSSCKNTGDDYEDSDVEDISAYLLSKNNAIEPRSFQNSRHPSTEQKQENATTI	780
XX	invention. These sequences were created by mutating the wild type Factor	Db	721	MTALKVSSCKNTGDDYEDSDVEDISAYLLSKNNAIEPRSFQNSRHPSTEQKQENATTI	780
XX	VIII:C coding sequence (see AWF1357) using mutagenic primers. The	QY	781	PENDIEKTDPAFHRTMPKIQNVSSDMLLRQSPTHGLSLSDLOEAKYETFSDDPS	840
XX	analogues comprise a native Factor VIII:C polypeptide modified at a site	Db	781	PENDIEKTDPAFHRTMPKIQNVSSDMLLRQSPTHGLSLSDLOEAKYETFSDDPS	840
XX	adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg	QY	841	PGADSNNSLSMTHFRPOLHSGDMVTPRSGLOLRNEKLGTTAATLKKLDFKVSST	900
XX	dipeptide is created. Factor VIII:C is a large glycoprotein that	Db	841	PGADSNNSLSMTHFRPOLHSGDMVTPRSGLOLRNEKLGTTAATLKKLDFKVSST	900
XX	participates in the blood coagulation cascade that ultimately converts	QY	901	NNLHISTIPSDNLAAGTDNTSSLGPPSPMVHYDSQDITTLFGKSSPLTESGGPLSLSEE	960
XX	soluble fibrinogen to insoluble fibrin clot effecting haemostasis. A	Db	901	NNLHISTIPSDNLAAGTDNTSSLGPPSPMVHYDSQDITTLFGKSSPLTESGGPLSLSEE	960
XX	deficiency in Factor VIII:C is responsible for haemophilia A, which is an	QY	961	NNDSKLLSGLMNSQESSGKNVSTESGRIFPKGKRAHGPALLTKDNALFKVSI SLLKTN	1020
XX	X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is	Db	961	NNDSKLLSGLMNSQESSGKNVSTESGRIFPKGKRAHGPALLTKDNALFKVSI SLLKTN	1020
XX	activated by plasma proteases, such as thrombin. During activation the	QY	1021	KTSNNSATNRKTHIDGPSILLIENSPSVWQNTLESDETFKVTLPILHDMKMDKNTALRL	1080
XX	mature polypeptide is cleaved to generate heavy and light chain fragments	Db	1021	KTSNNSATNRKTHIDGPSILLIENSPSVWQNTLESDETFKVTLPILHDMKMDKNTALRL	1080
XX	that are further cleaved. Complexes of two or more of the analogues,	QY	1081	NHMSNKTSSKNMNVQKKGPIPPDAQNPDMSFFKMLFLPESARWQRTKGNKNSLNG	1140
XX	nucleic acids and vectors encoding them may be used alone or in	Db	1081	NHMSNKTSSKNMNVQKKGPIPPDAQNPDMSFFKMLFLPESARWQRTKGNKNSLNG	1140
XX	conjunction with each other, for the prevention or treatment of active	QY	1141	QGPSPKQLVSLGPEKSVBQGNFLSKNVKVGKGFVKDVLKEMVFPSSRNLFITNLN	1200
XX	Factor VIII:C deficiency in a mammal. The analogues may be used as	Db	1141	QGPSPKQLVSLGPEKSVBQGNFLSKNVKVGKGFVKDVLKEMVFPSSRNLFITNLN	1200
XX	immunogens to raise antibodies, and in the treatment of haemophiliacs, by	QY	1201	LHNNYTHQEKILQBEIEKKETLIQENTVLPQIHVTVTGTFKNFMKNLFLILSTRQNVESYD	1260
XX	improvement of haemostasis. The analogues are resistant to proteolytic	Db			
XX	cleavage and display increased plasma half-life. They may be administered				
XX	at lower dosages and by different modes of administration				
XX	Sequence 2351 AA;				
SQ	Query Match 99.9%; Score 12405; DB 2; Length 2351;				
	Best Local Similarity 99.9%; Pred. No. 0;				
	Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	1 MQEISTCFLLRCFNSATRYILGAVELSDYMQSDLGELPVDARPPVPKSPFPN 60				
Db	1 MQEISTCFLLRCFNSATRYILGAVELSDYMQSDLGELPVDARPPVPKSPFPN 60				
QY	61 TSVVYKTLFVEVTDHLFNIAKPRPPWMLLGTPTQAEVVDVTVITLKNMASHPVSLHAV 120				
Db	61 TSVVYKTLFVEVTDHLFNIAKPRPPWMLLGTPTQAEVVDVTVITLKNMASHPVSLHAV 120				

Db 1201 LHENNTNQEKKIOEELEKKEETLLQSNVVLPIQHTVTGTFKFMKNLFLLSIRQNVESYD 1260
Qy 1261 GAYAPVLQDFASLNDSTNRKHTAHFSKKGHEENLEGLGNQTKQIVKCYACTTRISPNT 1320
Db 1261 GAYAPVLQDFASLNDSTNRKHTAHFSKKGHEENLEGLGNQTKQIVKCYACTTRISPNT 1320
Qy 1321 SQQNFTQSRKALQKQRLPLEEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFTQSRKALQKQRLPLEEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGAITQSPSLCLTRSHSIPQANSRPLPIAKVSPSPSIRPIYLRVLFPQDNSSHLPAASY 1440
Db 1381 KGAITQSPSLCLTRSHSIPQANSRPLPIAKVSPSPSIRPIYLRVLFPQDNSSHLPAASY 1440
Qy 1441 RKDSGVQSRSHFTQGAKKNNLSAILTLEMTGQREVSGISGTSATNSVTYKKVENTVLP 1500
Db 1441 RKDSGVQSRSHFTQGAKKNNLSAILTLEMTGQREVSGISGTSATNSVTYKKVENTVLP 1500
Qy 1501 KPDLPKTSQKVELLPKVHIYOKDLPPTTSNGSPGHLDLVSGLLQTEGALKWNEANRP 1560
Db 1501 KPDLPKTSQKVELLPKVHIYOKDLPPTTSNGSPGHLDLVSGLLQTEGALKWNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWSQESPEKTAFFKKOTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWSQESPEKTAFFKKOTILSL 1620
Qy 1621 NACSNHAIAINEGONKPBIEVWAKGGRTERLCSQNPVVLKXHQREITRITLOSQOE 1680
Db 1621 NACSNHAIAINEGONKPBIEVWAKGGRTERLCSQNPVVLKXHQREITRITLOSQOE 1680
Qy 1681 IDYDDTISVEMKKEFDIYDEENOSPSFOKTRHYFIAAVERLWDMGSSPHVLNRP 1740
Db 1681 IDYDDTISVEMKKEFDIYDEENOSPSFOKTRHYFIAAVERLWDMGSSPHVLNRP 1740
Qy 1741 AQSGSVPOFKVVFQFTDGGSTQPLRGELNEHLGLLGPYIRAEVDENIMVTFPNQASR 1800
Db 1741 AQSGSVPOFKVVFQFTDGGSTQPLRGELNEHLGLLGPYIRAEVDENIMVTFPNQASR 1800
Qy 1801 PYSFYSSLI SYEEDORQAEPRKFNKTYFWKQVHHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSSLI SYEEDORQAEPRKFNKTYFWKQVHHMAPTKDEDFCKAWAYFSDV 1860
Qy 1861 DLEKDVHSLGLLGLVCHTNTLNPAGHQVTVQEFALPFTTFDETKSWYFTENMBNCR 1920
Db 1861 DLEKDVHSLGLLGLVCHTNTLNPAGHQVTVQEFALPFTTFDETKSWYFTENMBNCR 1920
Qy 1921 PCNIQMEDPTPKENVRFRHAINGYIMDTLPGLVNAQDQRIWYLLSMGSGNENIHSIFSGH 1980
Db 1921 PCNIQMEDPTPKENVRFRHAINGYIMDTLPGLVNAQDQRIWYLLSMGSGNENIHSIFSGH 1980
Qy 1981 VFTVBEKKEEYKVALYNLYPGVETVEMLPSKAGIWEVCECLIGEHLHAGNLTFLVYSNKC 2040
Db 1981 VFTVBEKKEEYKVALYNLYPGVETVEMLPSKAGIWEVCECLIGEHLHAGNLTFLVYSNKC 2040
Qy 2041 QTPLGWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPLGWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Qy 2101 HGITQGARQKSSYIISOPIIMYSLDGKKWQTYRGNSTGTLMWFFGNVDSSGIRKHNIPN 2160
Db 2101 HGITQGARQKSSYIISOPIIMYSLDGKKWQTYRGNSTGTLMWFFGNVDSSGIRKHNIPN 2160
Qy 2161 PPIIARYIRLHPTHYSIRSTLRWELMGCDLNSCPLGMSKAISSAQITASSYFTNMPA 2220
Db 2161 PPIIARYIRLHPTHYSIRSTLRWELMGCDLNSCPLGMSKAISSAQITASSYFTNMPA 2220
Qy 2221 TWSPSKARLHLOGRSNAPVQVNNPKELQVDFQKTMKVTGTTQGVKSLLSYVYKPEL 2280
Db 2221 TWSPSKARLHLOGRSNAPVQVNNPKELQVDFQKTMKVTGTTQGVKSLLSYVYKPEL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPLLTRYLRIHPQSWWHQIALRM 2340

Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPLLTRYLRIHPQSWWHQIALRM 2340
Qy 2341 EVLGCERQDDY 2351
Db 2341 EVLGCERQDDY 2351
RESULT 11
AAW11398
ID AAW11398 standard; protein; 2351 AA.
XX AAW11398;
AC AAW11398;
DT 18-NOV-1997 (first entry)
XX Active Factor VIII:C analogue, delta 746, + residue 746 insertion.
DE Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
XX fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;
KW proteolytic cleavage.
XX Homo sapiens.
OS Synthetic.
XX
PH Key
FT Peptide
FT Protein
FT Region
FT Domain
FT Misc-difference
FT Modified-site
FT Region
FT
XX WC9703195-A1.
XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR) CHIRON CORP.
XX Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
XX
PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophiliacs, by improvement of haemostasis.
XX Claim 23; Page; 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,

CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX

SQ Sequence 2351 AA;

Query Match 99.9%; Score 12405; DS 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOELSTQFFLCILRCPFSATRYVILGAVELSDYMQSDGLGELPVDARPPRPVPSKFPFN 60
DB 1 MOELSTQFFLCILRCPFSATRYVILGAVELSDYMQSDGLGELPVDARPPRPVPSKFPFN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLIGPTTCAEYVDVTWITLKNMASHPSLHAY 120
DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPMGLIGPTTCAEYVDVTWITLKNMASHPSLHAY 120
QY 121 GVSYWKASGEAYDDQTSQREKEDDKVPFGGSHYYVQVLYKENGPMASDPLCLTYSLSH 180
DB 121 GVSYWKASGEAYDDQTSQREKEDDKVPFGGSHYYVQVLYKENGPMASDPLCLTYSLSH 180
QY 181 VDLVXDLSGLIGALLVCEGSLAKKTQTLHKFILLFAVDEGKSWHSETNSIMOODR 240
DB 181 VDLVXDLSGLIGALLVCEGSLAKKTQTLHKFILLFAVDEGKSWHSETNSIMOODR 240
QY 241 AASAAWPKXHTVGVNVRSLPGLIGCHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNH 300
DB 241 AASAAWPKXHTVGVNVRSLPGLIGCHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISITELTAQTLMDLQGFLLSCHISHOHDGMEAYVKVDSCPEEPQLRMKNE 360
DB 301 ROASLEISITELTAQTLMDLQGFLLSCHISHOHDGMEAYVKVDSCPEEPQLRMKNE 360
QY 361 EAEDYDDDLTDSMDVVRDNDNSPSFIQIRSVAKXHPKTWHYIAAEBEDWDYAPLVLA 420
DB 361 EAEDYDDDLTDSMDVVRDNDNSPSFIQIRSVAKXHPKTWHYIAAEBEDWDYAPLVLA 420
QY 421 PDRSYKSVLNGPQIRGKYKVRPMAYDTDFXTREAIQHSIGLPLLYCEVGDIL 480
DB 421 PDRSYKSVLNGPQIRGKYKVRPMAYDTDFXTREAIQHSIGLPLLYCEVGDIL 480
QY 481 LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPLPGEIFYKXWTVTVEDGP 540
DB 481 LIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPLPGEIFYKXWTVTVEDGP 540
QY 541 TKSDPCLTRYSSFVNMRDLASGLIGPLICYKESVDQGNQIYSDXENVILPSVEDE 600
DB 541 TKSDPCLTRYSSFVNMRDLASGLIGPLICYKESVDQGNQIYSDXENVILPSVEDE 600
QY 601 NRSWYLTEIQRPLNPAGVQLSDPBFQASNIHMSINGYVFDLSQLSVCLHEVAYWYILS 660
DB 601 NRSWYLTEIQRPLNPAGVQLSDPBFQASNIHMSINGYVFDLSQLSVCLHEVAYWYILS 660
QY 661 IGACTQFLSVFPGGYTPKHNVHTDILTPPFGSEIVFMSMENPGLWILGCHNSDFNRG 720
DB 661 IGACTQFLSVFPGGYTPKHNVHTDILTPPFGSEIVFMSMENPGLWILGCHNSDFNRG 720
QY 721 MTALLKVSSCDKNTGDIYEDSYSDISAYLLSKNATBPRFSQNSRHPSTRQGFNATTI 780
DB 721 MTALLKVSSCDKNTGDIYEDSYSDISAYLLSKNATBPRFSQNSRHPSTRQGFNATTI 780
QY 781 PENDIEKTDWFAHRTPMPIKQVSSSLLMLLRQSPTPHGLSLDLQEAQYTFSDDPDS 840
DB 781 PENDIEKTDWFAHRTPMPIKQVSSSLLMLLRQSPTPHGLSLDLQEAQYTFSDDPDS 840
QY 841 PGAIIDNSNISLSEMTHPQLHSGDMVFTPASGLQLRLNEKLGTTAATLKLKLDKFKVSS 900
DB 841 PGAIIDNSNISLSEMTHPQLHSGDMVFTPASGLQLRLNEKLGTTAATLKLKLDKFKVSS 900

QY 901 SNNLSTIPSDNLAAGTNTSSILGPPSPMEVHYDSQDLDITLFGKKSSPLTESGGPLSLEE 960
DB 901 SNNLSTIPSDNLAAGTNTSSILGPPSPMEVHYDSQDLDITLFGKKSSPLTESGGPLSLEE 960
QY 961 NNDKLLFESGLMNSQESSWGKNVSTESGRLFKGRAGHPALLTKDNALFKVSISSLKTN 1020
DB 961 NNDKLLFESGLMNSQESSWGKNVSTESGRLFKGRAGHPALLTKDNALFKVSISSLKTN 1020
QY 1021 KTSNNSATNRKTHIDGSPSLLIENSFVQNLLESDETEPKKVTPLIHDRMLKDKATARL 1080
DB 1021 KTSNNSATNRKTHIDGSPSLLIENSFVQNLLESDETEPKKVTPLIHDRMLKDKATARL 1080
QY 1081 NMSNKTTSKXNMVQOKKEGPIPPDAQNPDMEFFKMLFPESARWQRTGKNSLNSG 1140
DB 1081 NMSNKTTSKXNMVQOKKEGPIPPDAQNPDMEFFKMLFPESARWQRTGKNSLNSG 1140
QY 1141 QGSPKQVLSLGPBKSVGQNFSLSEKKNVVGKGEFTKDVGLKMWFPSSNLFUTNLDN 1200
DB 1141 QGSPKQVLSLGPBKSVGQNFSLSEKKNVVGKGEFTKDVGLKMWFPSSNLFUTNLDN 1200
QY 1201 LHENNTNHOEKKIQEEIEKKETLLOENVVLQIHTVTCTKNPMKNLFLLSRQNVGSD 1260
DB 1201 LHENNTNHOEKKIQEEIEKKETLLOENVVLQIHTVTCTKNPMKNLFLLSRQNVGSD 1260
QY 1261 GAYAPVLODFRSLNDSTNRKTKHTAHFSSKGBEENLEGLNQTKQIVKCYACTTRISPT 1320
DB 1261 GAYAPVLODFRSLNDSTNRKTKHTAHFSSKGBEENLEGLNQTKQIVKCYACTTRISPT 1320
QY 1321 SQONFVTRSRKALQKPLPLETELEKRIIVDDTSTOWSKNMKHLTSTLTQIDYNEKE 1380
DB 1321 SQONFVTRSRKALQKPLPLETELEKRIIVDDTSTOWSKNMKHLTSTLTQIDYNEKE 1380
QY 1381 KCAITQSPSLDCLTPSHSIPQANRSPPLPIAKVSSPPSIRPIYLTIRVLFDNSHHPAASY 1440
DB 1381 KCAITQSPSLDCLTPSHSIPQANRSPPLPIAKVSSPPSIRPIYLTIRVLFDNSHHPAASY 1440
QY 1441 RKDGSVQESSHFLGAKKNNLSLAILTEMTGDQREVGLSGLTSATNSVYTKKVENTVLP 1500
DB 1441 RKDGSVQESSHFLGAKKNNLSLAILTEMTGDQREVGLSGLTSATNSVYTKKVENTVLP 1500
QY 1501 KPDLPTSGKVLLPKVHIYQKDLPTFTSNGSPGHLDLVREGSLQGTGALIKWNEANRP 1560
DB 1501 KPDLPTSGKVLLPKVHIYQKDLPTFTSNGSPGHLDLVREGSLQGTGALIKWNEANRP 1560
QY 1561 GKVPFRVATSSAKTPSKLLDPLAWDNHYGTQIPKEWKSOEKSPEKTAFFKKDTILSL 1620
DB 1561 GKVPFRVATSSAKTPSKLLDPLAWDNHYGTQIPKEWKSOEKSPEKTAFFKKDTILSL 1620
QY 1621 NACESNHAAINEGONKPEIEVTWAKQGRTERLCSQNPVILKRHQRETRITTLQSDQEE 1680
DB 1621 NACESNHAAINEGONKPEIEVTWAKQGRTERLCSQNPVILKRHQRETRITTLQSDQEE 1680
QY 1681 IDYDDTISVEMKEDFDIYDEBENQSPRSFQKTRHYFIAAVERLWDYQSSPHVLRN 1740
DB 1681 IDYDDTISVEMKEDFDIYDEBENQSPRSFQKTRHYFIAAVERLWDYQSSPHVLRN 1740
QY 1741 AOSGSVPQKVVQFQFTDGSFTQPLYRGELNEHLGLGPIYRAEVEDNIMWTFRQASR 1800
DB 1741 AOSGSVPQKVVQFQFTDGSFTQPLYRGELNEHLGLGPIYRAEVEDNIMWTFRQASR 1800
QY 1801 PYSFYSSLISYEDDQGAEPKRNFKVKNETKITYPWKVQHMAPTKDEPCKAWAYFSDV 1860
DB 1801 PYSFYSSLISYEDDQGAEPKRNFKVKNETKITYPWKVQHMAPTKDEPCKAWAYFSDV 1860
QY 1861 DLEKQVHSLIGLPELLVCHTNTLNPAHGRQVTVQEFALFTIIFDETKSWYFTENMEENCA 1920
DB 1861 DLEKQVHSLIGLPELLVCHTNTLNPAHGRQVTVQEFALFTIIFDETKSWYFTENMEENCA 1920
QY 1921 PCNIQMEDTFKENREFAHNGVIMDTLPGVMAQDQRIWRVYLLSMGNSNENHSHFSGH 1980
DB 1921 PCNIQMEDTFKENREFAHNGVIMDTLPGVMAQDQRIWRVYLLSMGNSNENHSHFSGH 1980

1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKC 2040
1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKC 2040
2041 QTPIGMASGHIRDFOITASQYGGWAPKLAHLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
2041 QTPIGMASGHIRDFOITASQYGGWAPKLAHLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
2101 HGKIQAROKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLWYFFGNDSSGKHNIE 2160
2101 HGKIQAROKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLWYFFGNDSSGKHNIE 2160
2161 PPIIARVIRLHPTYSIRSTRIRLMEIMGCDLNSCMPGMSKALSDAQITASSYFTNFA 2220
2161 PPIIARVIRLHPTYSIRSTRIRLMEIMGCDLNSCMPGMSKALSDAQITASSYFTNFA 2220
2221 TWSPSKARLHLQGRSNAWRPQNNPKWLQVDFQTKMKVTQVTTQGVKSLTSMYVKEFL 2280
2221 TWSPSKARLHLQGRSNAWRPQNNPKWLQVDFQTKMKVTQVTTQGVKSLTSMYVKEFL 2280
2281 ISSSQDGHQWTLRFQNGKVKVFOGNDSPFTPVNSLOPPLTRYLRTHPOSVHQAIALRM 2340
2281 ISSSQDGHQWTLRFQNGKVKVFOGNDSPFTPVNSLOPPLTRYLRTHPOSVHQAIALRM 2340
2341 EVLGCEAQDLY 2351
2341 EVLGCEAQDLY 2351

RESULT 12

AAW11425
ID AAW11425 standard; protein; 2351 AA.
XX AAW11425;
XX AAW11425;
DT 20-NOV-1987 (first entry)
DE Active Factor VIII:C analogue T1308X.
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW proteolytic cleavage.
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT /note= "signal peptide"
FT Protein 20..2351
FT /note= "mature Factor VIII:C"
FT Region 20..1667
FT /note= "heavy chain fragment"
FT Domain 760..1667
FT /note= "B domain"
FT Modified-site 1327
FT /label= Phe, Glu, Pro
FT Region 1668..2350
FT /note= "light chain fragment"
XX
PN WO9703195-A1.
XX
PD 30-JAN-1997.
XX
PF 09-JUL-1996; 96WO-US011444.
XX
PR 11-JUL-1995; 95US-0001025P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Fung DT, Cohen FE, Innis M;
XX

DR WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
PT used in the treatment of haemophiliacs, by improvement of haemostasis.
XX Claim 28; Page; 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Bro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
SQ Sequence 2351 AA;
Query Match 99.9%; Score 12405; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOIELSTCFE-CLLRFCFSATRYVLGAVELSDWYDMSDLGELPVDARFPRVPKSPFN 60
DB 1 MOIELSTCFE-CLLRFCFSATRYVLGAVELSDWYDMSDLGELPVDARFPRVPKSPFN 60
QY 61 TSVVYKTLFVEFTDHLFNIAPRPPMMGLLGPTIQAEVYDVTITLKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAPRPPMMGLLGPTIQAEVYDVTITLKNMASHPVSLHAV 120
QY 121 GVSVMKASEGAEYDDQTSQREKDDKVPFGGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
DB 121 GVSVMKASEGAEYDDQTSQREKDDKVPFGGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKPILLFAVDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKPILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASAPAWPKMHTVNGYVNRSLPGLIGCHRXSVYVHVIKMGTTPEVHSIFLEGHTFLVRNH 300
DB 241 AASAPAWPKMHTVNGYVNRSLPGLIGCHRXSVYVHVIKMGTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLMLDGLQFLLSCHSHSHODGMEAVKVDSPESQLEMKNNE 360
DB 301 ROASLEISPIITFLTAQTLMLDGLQFLLSCHSHSHODGMEAVKVDSPESQLEMKNNE 360
QY 361 EAEDYDDDLTDSEMDVVRFDNDNSPFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLA 420
DB 361 EAEDYDDDLTDSEMDVVRFDNDNSPFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVLA 420
QY 421 PDDRSYKSYQLANGPQRIGRYKRVFMAYTDETFKTREAIQHSIGILGPLLYGVGDTL 480
DB 421 PDDRSYKSYQLANGPQRIGRYKRVFMAYTDETFKTREAIQHSIGILGPLLYGVGDTL 480
QY 481 LIIFKNQASRPYNTYPHGITDVRPLYSRLPKGVHLKDFPILPGEI PKYKWTVTVEQGP 540
DB 481 LIIFKNQASRPYNTYPHGITDVRPLYSRLPKGVHLKDFPILPGEI PKYKWTVTVEQGP 540
QY 541 TKSDPRCLTRYVSSFWNMERDLASGLIGPLLCYKESVDQGNQIMSKRNVILSVFDE 600
DB 541 TKSDPRCLTRYVSSFWNMERDLASGLIGPLLCYKESVDQGNQIMSKRNVILSVFDE 600

QY 601 NRSYLIENITQRPENPACVOLPEBPQASINMHSINGYVPSLQSLVCLHEVAVWYLS 660
DB 601 NRSYLIENITQRPENPACVOLPEBPQASINMHSINGYVPSLQSLVCLHEVAVWYLS 660
QY 661 IGAGTDFLSVFFSGYTFKHKVYEDTTLTLPFSGETVFMSENPGMLILGCHNSDFNRG 720
DB 661 IGAGTDFLSVFFSGYTFKHKVYEDTTLTLPFSGETVFMSENPGMLILGCHNSDFNRG 720
QY 721 MTALLKVSCKNCDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
DB 721 MTALLKVSCKNCDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
QY 781 PENDIEKTDPMFAHRTMPKIQNVSSDLMMLLQSPHPGLSLSDQEAKEYEFSDDPS 840
DB 781 PENDIEKTDPMFAHRTMPKIQNVSSDLMMLLQSPHPGLSLSDQEAKEYEFSDDPS 840
QY 841 PGAIDSNLSGEMTHFRPQLHSGDMVTPPSGLQLRLNEKLGTTAATLKKLDFKVSST 900
DB 841 PGAIDSNLSGEMTHFRPQLHSGDMVTPPSGLQLRLNEKLGTTAATLKKLDFKVSST 900
QY 901 SNNLITIPSNLAAGTDNTSSLOPPSPVHYDSQDITTLFGKXSSPLTESGGPLSSEE 960
DB 901 SNNLITIPSNLAAGTDNTSSLOPPSPVHYDSQDITTLFGKXSSPLTESGGPLSSEE 960
QY 961 NNDKSLLESGLMNSQESSGKNVSTESGRFLFKGRAHPALLTKDNALFKVSTISLLKTN 1020
DB 961 NNDKSLLESGLMNSQESSGKNVSTESGRFLFKGRAHPALLTKDNALFKVSTISLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSILLIENGSPVWQNTLESDETEFKVTPLIHDMMLDKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGPSILLIENGSPVWQNTLESDETEFKVTPLIHDMMLDKNATALRL 1080
QY 1081 NEMSKNTSSKNMVMQKKGPIPPDAQNQDMSFFKMLFPESARWQRTGKNSLNSG 1140
DB 1081 NEMSKNTSSKNMVMQKKGPIPPDAQNQDMSFFKMLFPESARWQRTGKNSLNSG 1140
QY 1141 QGPPKQVLVSLGPEKSVGQNFPLSKNVVGVKGFTKDVGKEMVFPSSNKLFLTNLDN 1200
DB 1141 QGPPKQVLVSLGPEKSVGQNFPLSKNVVGVKGFTKDVGKEMVFPSSNKLFLTNLDN 1200
QY 1201 LHENNTNHOEKIIOEIEPKETLQENNVLPQIHVTGTQKPMKMLFLLSTRQNVGSYD 1260
DB 1201 LHENNTNHOEKIIOEIEPKETLQENNVLPQIHVTGTQKPMKMLFLLSTRQNVGSYD 1260
QY 1261 GAYAPVLODFRSLNDSTNRKKTAAHPSKKGEEENLEGIGNTKQIVKVIACITRISPT 1320
DB 1261 GAYAPVLODFRSLNDSTNRKKTAAHPSKKGEEENLEGIGNTKQIVKVIACITRISPT 1320
QY 1321 SQCNFVTOQRSEALKQFRLPLETELEKRIIVDDTSTQMSKNMKHLTPSTLQIDYNEKE 1380
DB 1321 SQCNFVTOQRSEALKQFRLPLETELEKRIIVDDTSTQMSKNMKHLTPSTLQIDYNEKE 1380
QY 1381 KGAIQTQSLSDCLTRSHSIPQANRPLPIAKVSSPSPRIPIYLVTRVLFDQNSHLPAAZY 1440
DB 1381 KGAIQTQSLSDCLTRSHSIPQANRPLPIAKVSSPSPRIPIYLVTRVLFDQNSHLPAAZY 1440
QY 1441 RKKDSGVQESSHFLQAGKNNLSALLTLEMGDQREVGLGTSATNSVTKKVENTVLP 1500
DB 1441 RKKDSGVQESSHFLQAGKNNLSALLTLEMGDQREVGLGTSATNSVTKKVENTVLP 1500
QY 1501 KPDLPKTSKVELLPKVHIYQKDLFPFTETSNPSGHLDLVEGSLQGTGEGAIKMWANRP 1560
DB 1501 KPDLPKTSKVELLPKVHIYQKDLFPFTETSNPSGHLDLVEGSLQGTGEGAIKMWANRP 1560
QY 1561 GKVPFLRVATSSAKTPSKLPLDPLANDNHYGTQIPKEEWKSEKSPKTAFFKXDTILSL 1620
DB 1561 GKVPFLRVATSSAKTPSKLPLDPLANDNHYGTQIPKEEWKSEKSPKTAFFKXDTILSL 1620
QY 1621 NACESNHAIAANEGQNKPEIEVWAKQRTSELCSQNPVLRKHOREITRITLQSDQEE 1680
DB 1621 NACESNHAIAANEGQNKPEIEVWAKQRTSELCSQNPVLRKHOREITRITLQSDQEE 1680

QY 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYKSSSPHVLRNR 1740
DB 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYKSSSPHVLRNR 1740
QY 1741 AQSGSVQPKKVPQBPFTDGSFTQPLYRGELNEHLGGLGYIRAEVEDNIXVTFRQASR 1800
DB 1741 AQSGSVQPKKVPQBPFTDGSFTQPLYRGELNEHLGGLGYIRAEVEDNIXVTFRQASR 1800
QY 1801 PYSFYSSLIISVEDORQAGPRKXNVKPNETKHYFWKVOHEMAFTKDEPFCKAWAYFSDV 1860
DB 1801 PYSFYSSLIISVEDORQAGPRKXNVKPNETKHYFWKVOHEMAFTKDEPFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGILGPLLCHTNTLPAHGRQVTVQEFALFTIFDSTKSWYFTEENMCR 1920
DB 1861 DLEKDVHSGILGPLLCHTNTLPAHGRQVTVQEFALFTIFDSTKSWYFTEENMCR 1920
QY 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQRIWYLLSMGSENENIHSIHFSGH 1980
DB 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQRIWYLLSMGSENENIHSIHFSGH 1980
QY 1981 VFTVRKKEBYKVALYNLYPGVFEIVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
DB 1981 VFTVRKKEBYKVALYNLYPGVFEIVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
QY 2041 QTPLGMASGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLGMASGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQAROKFSSLIYSOFIIMYSLDGKKWQTVRGNSTGTLMVFPFGNVDSGIGKHIFN 2160
DB 2101 HGIKTQAROKFSSLIYSOFIIMYSLDGKKWQTVRGNSTGTLMVFPFGNVDSGIGKHIFN 2160
QY 2161 PPIIARVIRLHPHTHYRSRSTLRMELMGCNDLNSCMPLGESKASISDAQITASSYFTNMPA 2220
DB 2161 PPIIARVIRLHPHTHYRSRSTLRMELMGCNDLNSCMPLGESKASISDAQITASSYFTNMPA 2220
QY 2221 TWSFKARLHLQGRSNARWQVNNPKWLQVDFPQKTMKVTVGTTQGVKSLITSYKVEFL 2280
DB 2221 TWSFKARLHLQGRSNARWQVNNPKWLQVDFPQKTMKVTVGTTQGVKSLITSYKVEFL 2280
QY 2281 ISSQDGHQWTLFPQNGKVKVFCQNGQSDFTPVVNSLDPPELLTRYLRHPQSWVHQIALRM 2340
DB 2281 ISSQDGHQWTLFPQNGKVKVFCQNGQSDFTPVVNSLDPPELLTRYLRHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
DB 2341 EVLGCEAQDLY 2351
RESULT 13
AAW11362
ID AAW11362 standard; protein; 2351 AA.
XX
AC AAW11362;
XX
DT 18-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L277X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
XX proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT Peptide /note= "signal peptide"
FT Protein /note= "2351"
FT /note= "mature Factor VIII:C"
FT Region /note= "20..1667"

FT /note= "heavy chain fragment"
 FT 236
 FT /label= Phe, Glu, Pro
 FT 760..1667
 FT /note= "B domain"
 FT 1668..2350
 FT /note= "light chain fragment"
 XX
 XX W09703195-A1.
 XX
 XX 30-JAN-1997.
 XX
 XX 09-JUL-1996; 96WO-US011444.
 XX
 XX 11-JUL-1995; 95US-0001025P.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Hung DT, Cohen PE, Innis M;
 XX
 XX WPI; 1997-119050/11.
 DR
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 FT used in the treatment of haemophilias, by improvement of haemostasis.
 PT
 XX
 XX Claim 14; Page; 90pp; English.
 PS
 XX
 XX AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC x-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. they may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;
 Query Match 99.9%; Score 12405; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQEILSTCFCLIFRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPRVPSKSPFFN 60
 DB 1 MQEILSTCFCLIFRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPRVPSKSPFFN 60
 QY 61 TSVVYKTLFVEFTDHLNENIAKRPDPWMLLGPPTQAEVYDTVITLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLNENIAKRPDPWMLLGPPTQAEVYDTVITLKNMASHPVSLHAV 120
 QY 121 GVSVMKASEGABYDDQTSQREKEDDKVPFGGSHTYVWVLKENGPMASDPLCLTYSLSH 180
 DB 121 GVSVMKASEGABYDDQTSQREKEDDKVPFGGSHTYVWVLKENGPMASDPLCLTYSLSH 180
 QY 161 VDLVKDLSGLIGALLVCREGSLAKEXTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
 DB 161 VDLVKDLSGLIGALLVCREGSLAKEXTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
 QY 241 AASARAWPKMHTVNGYVNRSLPGLICGHRKSYVHWVINGTTPPEVHSIFLGEFTFLVRNH 300
 DB 241 AASARAWPKMHTVNGYVNRSLPGLICGHRKSYVHWVINGTTPPEVHSIFLGEFTFLVRNH 300

QY 301 ROASLEISPIITFLTAOTLLMDLGOPLLSCHISSHOHDGMEAYVKVDSCPEEPQOLRMKNE 360
 DB 301 ROASLEISPIITFLTAOTLLMDLGOPLLSCHISSHOHDGMEAYVKVDSCPEEPQOLRMKNE 360
 QY 361 EAEYDDEDLTDEMDVVRFDNDNSPSFIQIRSAKHPKTHVHYIAAEEEDWDYAPLVLA 420
 DB 361 EAEYDDEDLTDEMDVVRFDNDNSPSFIQIRSAKHPKTHVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDRSYKSQYLNNGPQIRGKYKVRPMAYTDETFKTREAIQHESGILGPLLYGEVGDTL 480
 DB 421 PDRSYKSQYLNNGPQIRGKYKVRPMAYTDETFKTREAIQHESGILGPLLYGEVGDTL 480
 QY 481 LIIFKNOASRPYNYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVTVDGP 540
 DB 481 LIIFKNOASRPYNYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVTVDGP 540
 QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQGNQIMSDKNVILFSVFDE 600
 DB 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQGNQIMSDKNVILFSVFDE 600
 QY 601 NRSWYLTENTQRELPNPAGVQLEDPEFQASNIHNSINGYVDFSLSQVCLHEVAYWVILS 660
 DB 601 NRSWYLTENTQRELPNPAGVQLEDPEFQASNIHNSINGYVDFSLSQVCLHEVAYWVILS 660
 QY 661 IGAQTDFLSVFFSGYTFKHMYEDTTLTPFSGETVPMGNENPGLMILGCHNSDFNRG 720
 DB 661 IGAQTDFLSVFFSGYTFKHMYEDTTLTPFSGETVPMGNENPGLMILGCHNSDFNRG 720
 QY 721 MTALLKVSSCKNTGDIYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 DB 721 MTALLKVSSCKNTGDIYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTDWFAHRTMPKIQNVSSDLMILRQSTPHGLSLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTDWFAHRTMPKIQNVSSDLMILRQSTPHGLSLSDLOEAKYETFSDDPS 840
 QY 841 PGALDSNNLSSEMTHFRPQLHSGDMVTFPSGQLRLNEKLGTTAATLKKLDFKVSST 900
 DB 841 PGALDSNNLSSEMTHFRPQLHSGDMVTFPSGQLRLNEKLGTTAATLKKLDFKVSST 900
 QY 901 SNNLIISTPDLNLAAGTNTSSLGPPSMVHYDSQDITTLFGKXSSPLTESGGPLSLEE 960
 DB 901 SNNLIISTPDLNLAAGTNTSSLGPPSMVHYDSQDITTLFGKXSSPLTESGGPLSLEE 960
 QY 961 NNDSKLLESGLMNSQESSWGKNSVSTESGLFKGKRAHGALLTKDNALFKVSI SLAKTN 1020
 DB 961 NNDSKLLESGLMNSQESSWGKNSVSTESGLFKGKRAHGALLTKDNALFKVSI SLAKTN 1020
 QY 1021 KTSNNSATNKRTHIDGPSLLIENSFVWQNILEDSTBFKVTPLIHDRMLMDKNATALRL 1080
 DB 1021 KTSNNSATNKRTHIDGPSLLIENSFVWQNILEDSTBFKVTPLIHDRMLMDKNATALRL 1080
 QY 1081 NHMKNKTTSSKNMNVQCKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
 DB 1081 NHMKNKTTSSKNMNVQCKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
 QY 1141 QGPSKQIVSLGPKSVGEGQNFLEKXKVVVGKEFTKDVGLKEMWFPSSNLFITNLN 1200
 DB 1141 QGPSKQIVSLGPKSVGEGQNFLEKXKVVVGKEFTKDVGLKEMWFPSSNLFITNLN 1200
 QY 1201 LHENNTNQEKKIOEEIEKKETLIQENWVLPQIHVGTGKNFMKNLFLSTQNVGEGSYD 1260
 DB 1201 LHENNTNQEKKIOEEIEKKETLIQENWVLPQIHVGTGKNFMKNLFLSTQNVGEGSYD 1260
 QY 1261 GAYAPVLDQFRLNDSNRTKGTFAHFSSKGBEENLEGLGNQTKQIVEKACTRIEISNT 1320
 DB 1261 GAYAPVLDQFRLNDSNRTKGTFAHFSSKGBEENLEGLGNQTKQIVEKACTRIEISNT 1320
 QY 1321 SQQNFVTCRSKRALQKQRLPLEETELSKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SQQNFVTCRSKRALQKQRLPLEETELSKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

```
QY 1381 KGATQSPSLDCLTRSHSIPOANESPLAKVSSFPSSIRPIYLTRVLFDQNSSHLPAASY 1440
DB 1381 KGATQSPSLDCLTRSHSIPOANESPLAKVSSFPSSIRPIYLTRVLFDQNSSHLPAASY 1440
QY 1441 RKDQSGVQESSHFFLQGAQKNNLSLAITLTLEMTGQREVSGISGTSATNSVTVKKVENTVLP 1500
DB 1441 RKDQSGVQESSHFFLQGAQKNNLSLAITLTLEMTGQREVSGISGTSATNSVTVKKVENTVLP 1500
QY 1501 KPDLPKTSKGVKVELLPKVHIYQKDLFPPTETNSGSGHLDLVEGSLLOQTEGAIKWNEANRP 1560
DB 1501 KPDLPKTSKGVKVELLPKVHIYQKDLFPPTETNSGSGHLDLVEGSLLOQTEGAIKWNEANRP 1560
QY 1561 GKVPFLRVATETSSAKTPSKLLDPLAWDNHYGTQTPKEWKSOEKSPEKTAFAKKDITLSL 1620
DB 1561 GKVPFLRVATETSSAKTPSKLLDPLAWDNHYGTQTPKEWKSOEKSPEKTAFAKKDITLSL 1620
QY 1621 NACSNHAIAINEGONKPEIEVWAKOGRTERLCQSNPPVLKRQHRIBITITLOSQDEE 1680
DB 1621 NACSNHAIAINEGONKPEIEVWAKOGRTERLCQSNPPVLKRQHRIBITITLOSQDEE 1680
QY 1681 IDYDTSIVEMKKEDFDIYDENQSPRSFQKTRHYTIAAVERLWDCMSSSPHVLNR 1740
DB 1681 IDYDTSIVEMKKEDFDIYDENQSPRSFQKTRHYTIAAVERLWDCMSSSPHVLNR 1740
QY 1741 AQSGSVPOFKVVOFEFTDGSFTQPLXRGELNEHLGLLGPVIRAEVEDNIMVTFRNQASR 1800
DB 1741 AQSGSVPOFKVVOFEFTDGSFTQPLXRGELNEHLGLLGPVIRAEVEDNIMVTFRNQASR 1800
QY 1801 PYSFYSSLISYEEDQQAEPKPNFVKNETKTYFWKQVQHNAPTKDFDCKAWAYFSDV 1860
DB 1801 PYSFYSSLISYEEDQQAEPKPNFVKNETKTYFWKQVQHNAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLIGPLVCHVTNTLNPAHGRQVTVQBEALPFTTIDETKSNVFTENYERNCA 1920
DB 1861 DLEKDVHSLIGPLVCHVTNTLNPAHGRQVTVQBEALPFTTIDETKSNVFTENYERNCA 1920
QY 1921 PCNIQMEPTFKENYRFAHNGIYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIFSGH 1980
DB 1921 PCNIQMEPTFKENYRFAHNGIYIMDTPLGLVMAQDQIRWYLLSGNSNENIHSIFSGH 1980
QY 1981 VFTVRKKEBKALVNLVPGVETVEMLPKSGAGIWRVECLIGELHLAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEBKALVNLVPGVETVEMLPKSGAGIWRVECLIGELHLAGMSTLFLVYSNKC 2040
QY 2041 QPLGVASGHIRDFOITASGQGVQWAPKLARLHYSGSINAWSTKPPSWIKVDLLAPMII 2100
DB 2041 QPLGVASGHIRDFOITASGQGVQWAPKLARLHYSGSINAWSTKPPSWIKVDLLAPMII 2100
QY 2101 HGKIQGARQKFSLSYISOFIIMYSLDGKKWQTYRGNSGTGLMWFFGNVDSSGIKHNIEN 2160
DB 2101 HGKIQGARQKFSLSYISOFIIMYSLDGKKWQTYRGNSGTGLMWFFGNVDSSGIKHNIEN 2160
QY 2161 PPIIARYIELHPTHYSIRSTLAMELNGCDLNSCSMPLGMESKAISDAQITASSYFTNMEA 2220
DB 2161 PPIIARYIELHPTHYSIRSTLAMELNGCDLNSCSMPLGMESKAISDAQITASSYFTNMEA 2220
QY 2221 TWSPSKARLHLOGRSNAPQVNNPKXEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
DB 2221 TWSPSKARLHLOGRSNAPQVNNPKXEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLTLYLRIHPQSWVHQIALRM 2340
DB 2281 ISSSQDGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPPLLTLYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
DB 2341 EVLGCEAQDLY 2351
```

RESULT 14

AAW11461

ID AAW11461 standard; protein; 2351 AA.

XX

```
AC AAW11461;
XX
XX 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue V171YK.
XX
XX Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;
XX proteolytic cleavage.
OS Homo sapiens.
XX Synthetic.
XX
XX Location/Qualifiers
XX Key 1..19
XX Peptide /note= "signal peptide"
XX Protein 20..2351
XX /note= "mature Factor VIII:C"
XX Region 20..1667
XX /note= "heavy chain fragment"
XX Domain 760..1667
XX /note= "B domain"
XX Region 1668..2350
XX /note= "light chain fragment"
XX Modified-site 1736
XX /label= Phe, Glu, Pro
XX
XX WO9703195-A1.
XX
XX 30-JAN-1997.
XX
XX 09-JUL-1996; 95WO-US011444.
XX
XX 11-JUL-1995; 95US-0001025P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Hung DT, Cohen PE, Innis M;
XX
XX WPI; 1997-119050/11.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophiliacs, by improvement of haemostasis.
XX
XX Claim 36; page; 90pp; English.
XX
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAR51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophiliacs, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;
```

Query Match

Best Local Similarity 99.9%; Score 12405; DB 2; Length 2351;

Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQ-ELSTCFCLLRFCFSATRYIYLCAVELSWDMQSDILGELPVDARFPRPVKSPFN 60
 Db 1 MQ-ELSTCFCLLRFCFSATRYIYLCAVELSWDMQSDILGELPVDARFPRPVKSPFN 60
 QY 61 TSVVYKTLFVEFTDHLFNTAKRPPMMGLLGPITIOAEVYDVTVITLKNWASHPVSLEAV 120
 Db 61 TSVVYKTLFVEFTDHLFNTAKRPPMMGLLGPITIOAEVYDVTVITLKNWASHPVSLEAV 120
 QY 121 GVSVKASEGAEDDQTSOREKEDDKVFPGGSHYVVMQVILKENGPNASDPLCLTYSYLSH 180
 Db 121 GVSVKASEGAEDDQTSOREKEDDKVFPGGSHYVVMQVILKENGPNASDPLCLTYSYLSH 180
 QY 181 VDLVKDINSGLIGALLVCRGSLAKETOTLHKELLFAPVDEGKSHSTKNSLMODRD 240
 Db 181 VDLVKDINSGLIGALLVCRGSLAKETOTLHKELLFAPVDEGKSHSTKNSLMODRD 240
 QY 241 AASARAWPKVHTVNGYVNRSLPGLIGCHRKSVMYEVIMGTTTPVHSGIFLEGHTFLVRNH 300
 Db 241 AASARAWPKVHTVNGYVNRSLPGLIGCHRKSVMYEVIMGTTTPVHSGIFLEGHTFLVRNH 300
 QY 301 ROASLEISPTFLTAOTLLMDLGOPLFSLCHSHQHDGMEAYVVKDSCPEBPQRMKONE 360
 Db 301 ROASLEISPTFLTAOTLLMDLGOPLFSLCHSHQHDGMEAYVVKDSCPEBPQRMKONE 360
 QY 420 EAEYDDDLTDSMDVRFDDNSPFTQBSVAKGPKTWVHIAAREEDWDYAPLVLA 420
 Db 420 EAEYDDDLTDSMDVRFDDNSPFTQBSVAKGPKTWVHIAAREEDWDYAPLVLA 420
 QY 421 PDDRSYKSOYLNNQPORIGRYKVRFMAYDFTFKTREAIOHESGILGPLLYGEVGDTL 480
 Db 421 PDDRSYKSOYLNNQPORIGRYKVRFMAYDFTFKTREAIOHESGILGPLLYGEVGDTL 480
 QY 481 LIIIPKQASRYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILGBIPKYKWTVTVEDGP 540
 Db 481 LIIIPKQASRYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILGBIPKYKWTVTVEDGP 540
 QY 541 TKSDPRCLTRYSSVNMERDIASLIGLPLICYKESVDQRGNQMSKRVNVLFSVFEDE 600
 Db 541 TKSDPRCLTRYSSVNMERDIASLIGLPLICYKESVDQRGNQMSKRVNVLFSVFEDE 600
 QY 601 NRSWLTENIORFLPNPAGVLEDEPFOASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
 Db 601 NRSWLTENIORFLPNPAGVLEDEPFOASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
 QY 661 IGAQTDFTSVFFSGYTFPKHMYEDTLTJFPFSGETVFMSENPLGILGCHNSDFNRG 720
 Db 661 IGAQTDFTSVFFSGYTFPKHMYEDTLTJFPFSGETVFMSENPLGILGCHNSDFNRG 720
 QY 721 MTALLKSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
 Db 721 MTALLKSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTDPPFAHRTMPKIQNVSSSOLLMLLRQSPTHGLSLSDLOEAKYETESDDPS 840
 Db 781 PENDIEKTDPPFAHRTMPKIQNVSSSOLLMLLRQSPTHGLSLSDLOEAKYETESDDPS 840
 QY 841 PGAIIDNNLSLSTHFRPOLHSGDMWFTPESGLOLRINEKLGTTAAATELKKLDFKVSST 900
 Db 841 PGAIIDNNLSLSTHFRPOLHSGDMWFTPESGLOLRINEKLGTTAAATELKKLDFKVSST 900
 QY 901 SNNLISTIPSDNLAAGTDNTSSIGPPSPVHVDSOLDTTLFGKKSPLTESGGPLSLGRE 960
 Db 901 SNNLISTIPSDNLAAGTDNTSSIGPPSPVHVDSOLDTTLFGKKSPLTESGGPLSLGRE 960
 QY 961 NNDSKLLSGLMNSQESSMGKNVSSSTESGRLPKGKAHPALLTKDNALFKVSISSLKTN 1020
 Db 961 NNDSKLLSGLMNSQESSMGKNVSSSTESGRLPKGKAHPALLTKDNALFKVSISSLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGFSLLIENSPPWQNIILSDTEPKKVTPLIHDRMLMDKNATLRL 1080
 Db 1021 KTSNNSATNRKTHIDGFSLLIENSPPWQNIILSDTEPKKVTPLIHDRMLMDKNATLRL 1080

QY 1081 NEMSNTTSSKNMEMVQOKKEGPIPDQAQNDMSFFQALFLPESARWIOQTHGKNLSNG 1140
 Db 1081 NEMSNTTSSKNMEMVQOKKEGPIPDQAQNDMSFFQALFLPESARWIOQTHGKNLSNG 1140
 QY 1141 QGSPKQLVSLGPEKSVEGQNFISEKXNVVVGGEFTKOVGLKEMVPPSSRNFLPLTNLDN 1200
 Db 1141 QGSPKQLVSLGPEKSVEGQNFISEKXNVVVGGEFTKOVGLKEMVPPSSRNFLPLTNLDN 1200
 QY 1201 LHENNTHQEKKEKTOBETTEKKTETLIOENVLPOJHTVTGTNKMKNLFLLLSTRQNVESYD 1260
 Db 1201 LHENNTHQEKKEKTOBETTEKKTETLIOENVLPOJHTVTGTNKMKNLFLLLSTRQNVESYD 1260
 QY 1261 GAYAPVLQDPRSLNDSNRTKXHTAHFSSKKEBENLEGLGNQTKQIIVEKYACTRISPT 1320
 Db 1261 GAYAPVLQDPRSLNDSNRTKXHTAHFSSKKEBENLEGLGNQTKQIIVEKYACTRISPT 1320
 QY 1321 SQQNFVQSRKALKQPLPLETELEBKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKE 1380
 Db 1321 SQQNFVQSRKALKQPLPLETELEBKRIIVDDTSTQWKNMKHLTPSLTQIDYNEKE 1380
 QY 1381 KGAITQSPDLCDLTRSHSIPQANESPLPIAKVSFPSPRIPIYLTREVLPQDNSSHLPAASY 1440
 Db 1381 KGAITQSPDLCDLTRSHSIPQANESPLPIAKVSFPSPRIPIYLTREVLPQDNSSHLPAASY 1440
 QY 1441 RKXDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
 Db 1441 RKXDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
 QY 1501 KPDLPKTSGKVELLPKHIIYQKOLFPPTETNSGSPGHLDLVEGSLLOGTEGAIKWEANRP 1560
 Db 1501 KPDLPKTSGKVELLPKHIIYQKOLFPPTETNSGSPGHLDLVEGSLLOGTEGAIKWEANRP 1560
 QY 1561 GKVPFLVATESSAKTPSKLLDPLANDNHYGTQIPKEEWSQESPKPTAPKKKDTILSL 1620
 Db 1561 GKVPFLVATESSAKTPSKLLDPLANDNHYGTQIPKEEWSQESPKPTAPKKKDTILSL 1620
 QY 1621 NACESNEHAIAINEGQNKPEIEVTWAKQGRTERLCSONPVLKXHQREITRTTLQSQOEE 1680
 Db 1621 NACESNEHAIAINEGQNKPEIEVTWAKQGRTERLCSONPVLKXHQREITRTTLQSQOEE 1680
 QY 1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSPOKTRHYPIAAVERLWDYGMSSSHXLRNR 1740
 Db 1681 IDYDDTISVEMKKEDFDIYDEDENQSPRSPOKTRHYPIAAVERLWDYGMSSSHXLRNR 1740
 QY 1741 AOSGSVPOFKKVVQEBFTDGSFTQPLVGELENEHLGLLGPVIRAEVEDNIIMVTRNOASR 1800
 Db 1741 AOSGSVPOFKKVVQEBFTDGSFTQPLVGELENEHLGLLGPVIRAEVEDNIIMVTRNOASR 1800
 QY 1801 PYSVYSSLSIYEEDQORQGAEPKPNVKNETKTYFWKVOHMAPTKDBFDCAKWAYPSDV 1860
 Db 1801 PYSVYSSLSIYEEDQORQGAEPKPNVKNETKTYFWKVOHMAPTKDBFDCAKWAYPSDV 1860
 QY 1861 DLEKDVHSGILGPLLCHTNTLNPAHGRQVTVQSPALFTTIFDETYSKYVFTENNERNCA 1920
 Db 1861 DLEKDVHSGILGPLLCHTNTLNPAHGRQVTVQSPALFTTIFDETYSKYVFTENNERNCA 1920
 QY 1921 PCNIOMEDPTFKENYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSMGSGNENIHSIHSGH 1980
 Db 1921 PCNIOMEDPTFKENYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSMGSGNENIHSIHSGH 1980
 QY 1981 VPTVRKKSEYKMALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSKNC 2040
 Db 1981 VPTVRKKSEYKMALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSKNC 2040
 QY 2041 QTPGLWASGHIRDTOITASGOYQWAPKLARLHVSYSINAMSTKEPPSWIKVDLLAPMII 2100
 Db 2041 QTPGLWASGHIRDTOITASGOYQWAPKLARLHVSYSINAMSTKEPPSWIKVDLLAPMII 2100
 QY 2101 HGIKTQARQKQFSSLYTSQFIIMYSLDGKKWOTYRGNSTGTIMVFFGVNVDSSGKHNIFN 2160
 Db 2101 HGIKTQARQKQFSSLYTSQFIIMYSLDGKKWOTYRGNSTGTIMVFFGVNVDSSGKHNIFN 2160
 QY 2161 PPILARYIRLHPHYTHYSIRSTLNMELMGDLNSCMLPQWESKAIQSDAQITASSYTNMPPA 2220

Db 2161 PPIIARIIRLEPTTHYSIRKSTIRLMELMCGDILNCSMPJGSEKASDAQITASSFTWFA 2220
Qy 2221 TWSPSKARLHLQGRSNAWRPQNNPKREWLOVDFOKTMKVGTGTTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLQGRSNAWRPQNNPKREWLOVDFOKTMKVGTGTTQGVKSLLTSMYVKEFL 2280
Qy 2281 ISSSQDGHOWTLFFONGKVKVFOGNDSTFPVNSLDPLLTRYLRIKHPQSWVHQIALRM 2340
Db 2281 ISSSQDGHOWTLFFONGKVKVFOGNDSTFPVNSLDPLLTRYLRIKHPQSWVHQIALRM 2340
Qy 2341 EVLGEAODLY 2351
Db 2341 EVLGEAODLY 2351
RESULT 15
AAW11387
ID AAW11387 standard; protein; 2351 AA.
XX
AC AAW11387;
XX
DT 18-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue N357X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
FH
FT Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "signal peptide"
FT Protein 20..2351
FT Protein /note= "mature Factor VIII:C"
FT Region 20..1667
FT Modified-site 376
FT Domain /label= Phe, Glu, Pro
FT Domain 760..1667
FT Region /note= "B domain"
FT Region 1668..2350
FT Region /note= "light chain fragment"
PN WO9703195-A1.
XX
PD 30-JAN-1997.
XX
PF 09-JUL-1996; 96WC-US011444.
XX
PR 11-JUL-1995; 95US-0001025P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Hung DT, Cohen FB, Innis M;
XX
PI WPI; 1997-119050/11.
XX
PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
PT used in the treatment of haemophilias, by improvement of haemostasis.
XX
XX Claim 20; Page; 90pp; English.
XX
CC AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that

CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
SQ Sequence 2351 AA;
Query Match 99.9%; Score 12405; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MQIELSTCFPLCLLRFCFSATRYVLGAVELSWDMQSLGELPVDARFPPRPVKSPPFN 60
Db 1 MQIELSTCFPLCLLRFCFSATRYVLGAVELSWDMQSLGELPVDARFPPRPVKSPPFN 60
Qy 61 TSVVYKKTFLFVEFTDHLFNIAKPRPPMGLLGPTIOAEVDTVTITLKNMASHPVSLHAV 120
Db 61 TSVVYKKTFLFVEFTDHLFN-AKPRPPMGLLGPTIOAEVDTVTITLKNMASHPVSLHAV 120
Qy 121 GVSVMKASEGAEYDDQTSQREKDDKVPFGSHYVWQVLKENGFMASDPLCLTYSLSH 180
Db 121 GVSVMKASEGAEYDDQTSQREKDDKVPFGSHYVWQVLKENGFMASDPLCLTYSLSH 180
Qy 181 VDLVKOLNSGLIGALLVCREGSLAKEKQTQTLHKFILLFAVDEGKSWHSEKNSLMQDRD 240
Db 181 VDLVKOLNSGLIGALLVCREGSLAKEKQTQTLHKFILLFAVDEGKSWHSEKNSLMQDRD 240
Qy 241 AASARAWPKMTVNGYVNRSLPGLIGCHRKSVYVHVIQMGTTPEVHSIFLGHTFLVRNH 300
Db 241 AASARAWPKMTVNGYVNRSLPGLIGCHRKSVYVHVIQMGTTPEVHSIFLGHTFLVRNH 300
Qy 301 ROASLEISFIFTLTAQTLLMDLGQFLLSCHISSHODGMEAYVKVDSCEPEPOLEMKNE 360
Db 301 ROASLEISFIFTLTAQTLLMDLGQFLLSCHISSHODGMEAYVKVDSCEPEPOLEMKNE 360
Qy 361 EADYDDDLTDSEMDVVFDDNDSFQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
Db 361 EADYDDDLTDSEMDVVFDDNDSFQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
Qy 421 PDRSYKSYLNNGPQIRGRYKVRPMAYTDETFKTRATQHESGILGPLLYGEVGTDL 480
Db 421 PDRSYKSYLNNGPQIRGRYKVRPMAYTDETFKTRATQHESGILGPLLYGEVGTDL 480
Qy 481 LIIFKNQASRPNTYPHGKITDVRPLYSRRLPKGVKHLKDFPLPGEIIFKYKWTVTVEDGP 540
Db 481 LIIFKNQASRPNTYPHGKITDVRPLYSRRLPKGVKHLKDFPLPGEIIFKYKWTVTVEDGP 540
Qy 541 TKSDPRCLTRYSSFWNMERDLASGLIGPLLI CYKESVDQQRGNQIMSDKRNVLFSVPDE 600
Db 541 TKSDPRCLTRYSSFWNMERDLASGLIGPLLI CYKESVDQQRGNQIMSDKRNVLFSVPDE 600
Qy 601 NRSWYLTEINIQRFLENPAGVQLEDPEFQASNTIMHSINGYVFDLSQLSVCLHVAWYILS 660
Db 601 NRSWYLTEINIQRFLENPAGVQLEDPEFQASNTIMHSINGYVFDLSQLSVCLHVAWYILS 660
Qy 661 IGAQTDLSVFPFSGYTFKHVMYVETLTLFPFSGETVFMSENGPGLWILGCHNSPFRNG 720
Db 661 IGAQTDLSVFPFSGYTFKHVMYVETLTLFPFSGETVFMSENGPGLWILGCHNSPFRNG 720
Qy 721 MTALLKVSSCDNTGYVEDSYEDISAYLLSKNNAIEPRFSQNRHSTROKQFNATTI 780
Db 721 MTALLKVSSCDNTGYVEDSYEDISAYLLSKNNAIEPRFSQNRHSTROKQFNATTI 780

Qy 781 PENDIEKTDWFAHRTWMPKIONVSSDLLMLLRQSPTPHGLSLSDLQEAKEYETFSDDPS 840
Db 781 PENDIEKTDWFAHRTWMPKIONVSSDLLMLLRQSPTPHGLSLSDLQEAKEYETFSDDPS 840
Qy 841 PGAI DSNNISLSEMTHTFPQLHSGDMVFTPBSGLQLRINEKLGTTAAATFELKLLDFKVSST 900
Db 841 PGAI DSNNISLSEMTHTFPQLHSGDMVFTPBSGLQLRINEKLGTTAAATFELKLLDFKVSST 900
Qy 901 SNNLISITPSDNLAAGTNTSSIGPPSPHVDQLDITLFGKKSSPLTSGGGLSSEE 960
Db 901 SNNLISITPSDNLAAGTNTSSIGPPSPHVDQLDITLFGKKSSPLTSGGGLSSEE 960
Qy 961 NND SKLLESGLMNSOESSGKNVSTESGRFLFKGRAHGPALLTKONALFKVSTSLKTN 1020
Db 961 NND SKLLESGLMNSOESSGKNVSTESGRFLFKGRAHGPALLTKONALFKVSTSLKTN 1020
Qy 1021 KTSNSATNRKTHIDGPSLLIENSPSVWQNILESDETEPKVTPLIHDRMLMDXNATALRL 1080
Db 1021 KTSNSATNRKTHIDGPSLLIENSPSVWQNILESDETEPKVTPLIHDRMLMDXNATALRL 1080
Qy 1081 NHMKNKTTSSKNMENVQOKSGPIPPDAQNPMGFFFKMLFLPESARWIQRTHGKNSLNSG 1140
Db 1081 NHMKNKTTSSKNMENVQOKSGPIPPDAQNPMGFFFKMLFLPESARWIQRTHGKNSLNSG 1140
Qy 1141 QGPSPKQIVSLGPEKSVBQNFLEKKNVWVGKGEFTKDVGLKEMVFPSSRNLFITNLND 1200
Db 1141 QGPSPKQIVSLGPEKSVBQNFLEKKNVWVGKGEFTKDVGLKEMVFPSSRNLFITNLND 1200
Qy 1201 LHENNTNQBKIKOEBIBKKTLLQENVVLQIHVITGTFKPMKMLFLLSTRQNWEGSYD 1260
Db 1201 LHENNTNQBKIKOEBIBKKTLLQENVVLQIHVITGTFKPMKMLFLLSTRQNWEGSYD 1260
Qy 1261 GAYAPVLQDPPSLNDSTNRKTHAHPSKGBEENLEGKQTKOIVBKVACTTISSENT 1320
Db 1261 GAYAPVLQDPPSLNDSTNRKTHAHPSKGBEENLEGKQTKOIVBKVACTTISSENT 1320
Qy 1321 SQQNFVTOQRKALQKRLPLEETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFVTOQRKALQKRLPLEETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGATQSPSLDCLTRSHSIPQANRSPDIPAKVSSPSPSRPIYLTRVLFDQNSSHLPAASY 1440
Db 1381 KGATQSPSLDCLTRSHSIPQANRSPDIPAKVSSPSPSRPIYLTRVLFDQNSSHLPAASY 1440
Qy 1441 RKDSGVQOESSEHTQAKKNLSLAILILEMTGDQ3VGSIGTSATNSVYKVENTVLP 1500
Db 1441 RKDSGVQOESSEHTQAKKNLSLAILILEMTGDQ3VGSIGTSATNSVYKVENTVLP 1500
Qy 1501 KPDLPKTSKGVVELLPKVHIYKDLPTTETSNGPSGHLDLVBGSLLOQTEGAIKMWENARP 1560
Db 1501 KPDLPKTSKGVVELLPKVHIYKDLPTTETSNGPSGHLDLVBGSLLOQTEGAIKMWENARP 1560
Qy 1561 GKVPPLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEBWSQEKSPKTAFAKKDITLSL 1620
Db 1561 GKVPPLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEBWSQEKSPKTAFAKKDITLSL 1620
Qy 1621 NACESNHAIAAINEGONKPEIEVTWAKQRTBRLCSQNPVVLKRHQREITRTTLOSQOEE 1680
Db 1621 NACESNHAIAAINEGONKPEIEVTWAKQRTBRLCSQNPVVLKRHQREITRTTLOSQOEE 1680
Qy 1681 IDYDDTISVEMKKEDFDYDEDENSPSPSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKKEDFDYDEDENSPSPSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Qy 1741 AOSGSVPQFKVVFQEFFTDGSGTQPLYRGELNEHGLGPIYRAEVEDNIMVTFRNQASR 1800
Db 1741 AOSGSVPQFKVVFQEFFTDGSGTQPLYRGELNEHGLGPIYRAEVEDNIMVTFRNQASR 1800
Qy 1801 PYSFYSSLSISYEDQOQAEPRKFNFKNETXTYFWKVOHMAPTKDBPDCKAWAYFSDV 1860
Db 1801 PYSFYSSLSISYEDQOQAEPRKFNFKNETXTYFWKVOHMAPTKDBPDCKAWAYFSDV 1860
Qy 1861 DLEKDVHESGLIGPLLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENNERCRA 1920

Db 1561 DLEKDVHESGLIGPLLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENNERCRA 1920
Qy 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVNAQDQRIWYLLSNGSNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGLVNAQDQRIWYLLSNGSNENIHSIHFSGH 1980
Qy 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
Db 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
Qy 2041 QTPLGMA SGHIRDFQITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPLGMA SGHIRDFQITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Qy 2101 HGIKTQAGAROKFSSLYISQFIIMYSIDGKKWQTVGKSTGTLMVFFGNDVSSGIKHNIEN 2160
Db 2101 HGIKTQAGAROKFSSLYISQFIIMYSIDGKKWQTVGKSTGTLMVFFGNDVSSGIKHNIEN 2160
Qy 2161 PPIIARVIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKASDAQITASSYFTNMFA 2220
Db 2161 PPIIARVIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKASDAQITASSYFTNMFA 2220
Qy 2221 TWSPSKARLHLQGRSNAWRPOVNNPKEWLQVDFOKTMKVITGVTTQGVKSLLTSMYKBEFL 2280
Db 2221 TWSPSKARLHLQGRSNAWRPOVNNPKEWLQVDFOKTMKVITGVTTQGVKSLLTSMYKBEFL 2280
Qy 2281 ISSSODGHQWTLFPONGKVKVFCGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Db 2281 ISSSODGHQWTLFPONGKVKVFCGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Qy 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

Search completed: April 13, 2004, 14:04:07
Job time : 77.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 27 Seconds
(without alignments)
8375.788 Million cell updates/sec

Title: NP000123-328
Perfect score: 12416
Sequence: 1 MQEISLTCFFLLRFCFSA.....WVHQIALMEVLGCEAQDIY 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: *
1: piri: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12410	100.0	2351	1 EZHU	coagulation factor
2	8827	71.1	2319	2 A47004	coagulation factor
3	8068	65.0	2133	2 T42783	coagulation factor
4	2823.5	22.7	2224	1 KFHU5	coagulation factor
5	2774.5	22.3	2183	2 T42764	coagulation factor
6	2747	22.1	2211	1 KFBOS	coagulation factor
7	2190	17.6	859	2 A25945	coagulation factor
8	1326	10.7	1069	1 KTHU	ferroxidase (EC 1.11.1.1)
9	1276	10.3	1059	1 A35210	ferroxidase (EC 1.11.1.1)
10	1104	8.9	216	2 A44258	factor VIII-associ
11	663	5.3	427	2 J04915	ags protein precu
12	657	5.3	463	1 T436479	milk fat globule m
13	650	5.2	409	2 T11743	pp47 protein - pig
14	635	5.1	401	2 S65138	Glycoprotein anti
15	635	5.1	427	2 S74211	PAS-677 protein pr
16	443	3.6	927	1 J00948	A5 antigen precurs
17	424.5	3.4	218	2 A47285	milk fat globule p
18	306.5	2.5	3133	2 S52093	hemocytin - silkw
19	262	2.1	845	2 JCS256	adipocyte transcri
20	239	1.9	1072	2 A86827	hypothetical prote
21	236.5	1.9	3418	1 G02334	breast cancer tumo
22	235.5	1.9	1358	2 A93360	SIR4 protein - yea
23	229	1.8	2166	2 G70163	hypothetical prote
24	224.5	1.8	3329	2 T30904	breast cancer tumo
25	223	1.8	2401	2 T28676	rhoctry protein -
26	223	1.8	2954	2 T14156	kinesin-related pr
27	222.5	1.8	3329	2 T42205	breast cancer susc
28	221	1.8	719	2 S51739	transcription repr
29	218.5	1.8	1928	2 S46773	myosin heavy chain

RESULT 1

EZHU
coagulation factor VIII precursor [validated] - human
N/Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant compo
C/Species: Homo sapiens (man)
C/Date: 28-Aug-1985 #sequence revision 28-Aug-1985 #text change 08-Dec-2000
C/Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; B42
R/Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A/Title: Sequence of the exon-containing regions of the human factor VIII gene.
A/Reference number: I54318; MUID:93265012; PMID:1303178
A/Accession: I54318
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1921, 'S', 1923-2351 <RES>
A/Cross-references: GB:M88648; NID:gi182381; PIDN:AAAS2420.1; PID:gi182383
R/Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seaburg,
Nature 312, 330-337, 1984
A/Title: Expression of active human factor VIII from recombinant DNA clones.
A/Reference number: A00525; MUID:85061548; PMID:6438526
A/Accession: A00525
A/Molecule type: mRNA
A/Residues: 1-2351 <WOO>
R/Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittmar, D.D.; I
S, D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A/Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A/Reference number: I58059; MUID:85061550; PMID:6438528
A/Accession: I58059
A/Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RE2>
A/Cross-references: GB:X01740; NID:gi182802; PIDN:AAAS2484.1; PID:gi182803
R/Truett, M.A.; Blacher, R.; Burke, D.; Chu, C.; Dina, D.; Hartog, K.; Kuo,
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.; No;
DNA 4, 333-349, 1985
A/Title: Characterization of the polypeptide composition of human factor VIII:C and the
A/Reference number: A23584; MUID:86081164; PMID:3935400
A/Accession: A23584
A/Molecule type: mRNA
A/Residues: 1-2351 <TRU>
A/Cross-references: GB:M14113; NID:gi182817; PIDN:AAAS2485.1; PID:gi182818
R/Eaton, D.; Rodriguez, G.B.; H.; Vehar, G.A.
Biochemistry 25, 505-512, 1986
A/Title: Proteolytic processing of human factor VIII. Correlation of specific cleavages
ity.
A/Reference number: A26174; MUID:86159740; PMID:3082357
A/Accession: A26174
A/Molecule type: protein
A/Residues: 20-361392-399, 'X', 401-402:1668-1678;1709-1722, 'D', 1723-1725;1741-1755 <EAT>
R/Pittmar, D.D.; Wang, J.H.; Kaufman, R.J.

ALIGNMENTS

hypothetical prote
probable calcium-b
hypothetical prote
hypothetical prote
tpr homolog - fru
breast cancer tumo
amiloride-sensitiv
hypothetical coile
ankyrin 2, neuro
neurexin IV - fru
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
similar to late em
vacuolar protein V

30 218.5 1.8 3507 2 T34513
31 217 1.7 1381 1 S45781
32 216.5 1.7 1271 2 D64237
33 214.5 1.7 5005 2 F82884
34 213.5 1.7 2346 2 T13829
35 213.5 1.7 3328 2 T30835
36 213 1.7 1420 1 A44361
37 212 1.7 1957 2 T38077
38 212 1.7 3924 2 S37431
39 211 1.7 1283 2 T13799
40 209.5 1.7 1177 2 I64233
41 208.5 1.7 1628 2 E90538
42 207.5 1.7 737 2 T31349
43 206.5 1.7 737 2 T15615
44 204.5 1.6 1825 2 T02712
45 204 1.6 1165 2 S62982

Biochemistry 31, 3315-3325, 1992
 A>Title: Identification and functional importance of tyrosine sulfate residues within re
 A:Reference number: A42348; PMID:92207952; PMID:1554716
 A:Accession: A42348
 A:Molecule type: protein
 A:Residues: 20-36/356-372/392-408/582-594/1669-1692/1693-1708/1709-17
 A:Experimental source: recombinant material from Chinese hamster ovary cells
 A>Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
 R:Pay, P.J.; Smudzins, T.M.
 J. Biol. Chem. 264, 14005-14010, 1999
 A>Title: Intersubunit fluorescence energy transfer in human factor VIII.
 A:Reference number: A43986; PMID:99340500; PMID:2503509
 A:Accession: A43986
 A:Molecule type: protein
 A:Residues: 'X', 517-523/1853-1860, 'X', 1862-1864, 'X', 1866 <P>
 R:Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;
 J. Biol. Chem. 266, 740-746, 1991
 A>Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
 A:Reference number: A56109; PMID:91033266; PMID:1898735
 A:Contents: annotation; sulfation
 R:Gitschier, J.; Wood, W.I.; Goralaka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
 Nature 312, 326-330, 1984
 A>Title: Characterization of the human factor VIII gene.
 A:Reference number: A56196; PMID:85061547; PMID:6438525
 A:Contents: annotation; introns
 R:McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
 Protein Sci. 4, 740-746, 1995
 A>Title: Locations of disulfide bonds and free cysteines in the heavy and light chains o
 A:Reference number: A56216; PMID:9533827; PMID:7613471
 A:Contents: annotation; disulfide bonds
 A>Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
 R:Kjalke, M.; Hedling, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
 Eur. J. Biochem. 234, 773-779, 1995
 A>Title: Amino acid residues 721-729 are required for full factor VIII activity.
 A:Reference number: S63527; PMID:96163459; PMID:8575434
 A:Accession: S63527
 A:Molecule type: protein
 A:Residues: 733-752/753-759 <K>
 R:Lind, P.; Larsson, K.; Spira, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg
 Eur. J. Biochem. 232, 19-27, 1995
 A>Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
 A:Reference number: S66445; PMID:96048024; PMID:7556150
 A:Accession: S66445
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1668-1685 <L>
 C:Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
 C:Genetics:
 A:Gene: GDB:F8C
 A:Cross-references: GDB:119124; OMIM:306700
 A:Map position: Xq28-Xq28
 A:Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
 A:Function:
 A>Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
 A:Pathway: Blood coagulation
 C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
 C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-235/Domain: coagulation factor VIII #status experimental <MAT>
 F:20-740/Product: coagulation factor VIIIa heavy chain #status experimental <ACH>
 F:20-356/Domain: A1 <DA1>
 F:232-748/Domain: ferroxidase repeat homology <FOL>
 F:392-759/Domain: A2 <DA2>
 F:402-730/Domain: ferroxidase repeat homology <FOL>
 F:760-1667/Domain: B <DBO>
 F:1668-2351/Product: coagulation factor VIIIa light chain #status experimental <ACL>
 F:1709-2038/Domain: A3 <DA3>
 F:1715-2038/Domain: ferroxidase repeat homology <FOL>
 F:2039-2191/Domain: C1 <DC1>
 F:2039-2188/Domain: discoidin I amino-terminal homology <DNI>
 F:2192-2351/Domain: C2 <DC2>
 F:2192-2345/Domain: discoidin I amino-terminal homology <DNI>
 F:2192-2345/Domain: discoidin I amino-terminal homology <DNI>
 F:60-258, 607, 776, 803, 847, 919, 962, 982, 1020, 1024, 1074, 1085, 1204, 1274, 1278, 1305, 1319, 1403, 1

F:172-138, 267-348, 547-573, 649-730, 1851-1877, 1918-1922, 2040-2188/Disulfide bonds: #status
 F:355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status predicted
 F:365, 737, 738, 742, 1683, 1699/Binding site: sulfate (Tyr, thrombin) #status experimental
 F:391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:414, 426/Binding site: sulfate (Tyr) (covalent) #status predicted
 F:753-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
 F:1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
 F:1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experimental
 F:2193-2345/Disulfide bonds: #status predicted
 Query Match 100.0%; Score 12410; DB 1; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MQIELSCFFCLILRPFCSATRRYYLGAVELSWDMQSDLGELFVDARPPRRPKSPFN 60
 Db 1 MQIELSCFFCLILRPFCSATRRYYLGAVELSWDMQSDLGELFVDARPPRRPKSPFN 60
 QY 61 TSVVVKTLFVEFTDHLFNIAPKPPPMGLLPPTIOAEVDTVVITLKNMASHPVSLHAV 120
 Db 61 TSVVVKTLFVEFTDHLFNIAPKPPPMGLLPPTIOAEVDTVVITLKNMASHPVSLHAV 120
 QY 121 GVSVMKASEGAEYDDQTSQREKEDDKVFPFGSSHTYVMQVLKENGPMASDPLCLTYSYLSH 180
 Db 121 GVSVMKASEGAEYDDQTSQREKEDDKVFPFGSSHTYVMQVLKENGPMASDPLCLTYSYLSH 180
 QY 181 VDLVKDLSNGHIGALLVCREGLAKETQTLHKPILLFAVDEGKSWHSEFNSLMQDRD 240
 Db 181 VDLVKDLSNGHIGALLVCREGLAKETQTLHKPILLFAVDEGKSWHSEFNSLMQDRD 240
 QY 241 AASARAPKMTVNGVYVNRSLPGLTGCXKSVVHVIQMTTPEVHSTFLEHGTPLVNRH 300
 Db 241 AASARAPKMTVNGVYVNRSLPGLTGCXKSVVHVIQMTTPEVHSTFLEHGTPLVNRH 300
 QY 301 ROASLEISPTFLTAQTLMDLQGLLSCHTSSHQHDMGMEAYVVDSCPEPQRLMKNE 360
 Db 301 ROASLEISPTFLTAQTLMDLQGLLSCHTSSHQHDMGMEAYVVDSCPEPQRLMKNE 360
 QY 361 EAEYDDDLTDSMDVVRFDNDNSPFIQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
 Db 361 EAEYDDDLTDSMDVVRFDNDNSPFIQIRSVAKKPKTWVHYIAAEEEDWDVAPLVLA 420
 QY 421 PDRSYKSOYLNNQPORIGRKVKVRFMAYDETFTKREAIQHSSEGLGLPYLGEVGDTL 480
 Db 421 PDRSYKSOYLNNQPORIGRKVKVRFMAYDETFTKREAIQHSSEGLGLPYLGEVGDTL 480
 QY 481 LIIFKNQASRPYNIYPHGIITVRPLYSRRLPKGVKHLKDPILPGEIPKYKWTVTVEDGP 540
 Db 481 LIIFKNQASRPYNIYPHGIITVRPLYSRRLPKGVKHLKDPILPGEIPKYKWTVTVEDGP 540
 QY 541 TKSDPRLCTRYSSFVNMERDLASGLIGPLLI CYKESVDQGNQIMSKRVILFSVFDE 600
 Db 541 TKSDPRLCTRYSSFVNMERDLASGLIGPLLI CYKESVDQGNQIMSKRVILFSVFDE 600
 QY 601 NRSWYLTENIQRFNPNAGVLEDPFQASNMHSINGYVDFSLQSLCYCLHEVAYWYLS 660
 Db 601 NRSWYLTENIQRFNPNAGVLEDPFQASNMHSINGYVDFSLQSLCYCLHEVAYWYLS 660
 QY 661 IGAQTDPLSVFSGVTFKHKVYEDTLTLPFGSETVFMSENFGMLWILGCHNDFNRG 720
 Db 661 IGAQTDPLSVFSGVTFKHKVYEDTLTLPFGSETVFMSENFGMLWILGCHNDFNRG 720
 QY 721 MTALLKVSSCDKNTGDDYEDSYEDISAYLLSKNNAIIPRSFSQNSRHPSTKQFNATTI 780
 Db 721 MTALLKVSSCDKNTGDDYEDSYEDISAYLLSKNNAIIPRSFSQNSRHPSTKQFNATTI 780
 QY 781 PENDIEKTDPMFAHRTMPKIQNVSSSDLLMLLRQSPFPHGLSLSDLOEAKYETFSDDPS 840
 Db 781 PENDIEKTDPMFAHRTMPKIQNVSSSDLLMLLRQSPFPHGLSLSDLOEAKYETFSDDPS 840
 QY 841 PCADSNNSLSEMTFRPQLHHSQDMVTFPSGLOLRNEKLGTTAATLTKLDFKVSST 900
 Db 841 PCADSNNSLSEMTFRPQLHHSQDMVTFPSGLOLRNEKLGTTAATLTKLDFKVSST 900

841 PGALDSNNLSLSEMTHTPQLHGHGDMVFTPESGQLRLNEKLTGTAATLTKLDFKVSST 900
901 SNHLISTIPSDNLAACTNDTSSLGPPSMPEHYDSQDITLFGKKSPLTDSGGPLSLEE 960
901 SNHLISTIPSDNLAACTNDTSSLGPPSMPEHYDSQDITLFGKKSPLTDSGGPLSLEE 960
961 NNDKSLLESGLMNSQESSWGKNVSTSGRLFKGKRAHGEPALLTKNALFKVSIILKTN 1020
961 NNDKSLLESGLMNSQESSWGKNVSTSGRLFKGKRAHGEPALLTKNALFKVSIILKTN 1020
1021 KTSNNSATNKTHIDGSPSLLIENSPPVQWNLSDTEFFKXVTPLIHDRLMDKNATAIRL 1080
1021 KTSNNSATNKTHIDGSPSLLIENSPPVQWNLSDTEFFKXVTPLIHDRLMDKNATAIRL 1080
1081 NMSNKTTSKQNMWVQCKEGPTPPDAONPDSPFFKWLFLPESARWIQTHGKNSLNSG 1140
1081 NMSNKTTSKQNMWVQCKEGPTPPDAONPDSPFFKWLFLPESARWIQTHGKNSLNSG 1140
1141 QGSPKQVSLGPEKSVGQNFSEKKNVVGKGEFTKDVGLKEMVPPSSRNPLFTNLDN 1200
1141 QGSPKQVSLGPEKSVGQNFSEKKNVVGKGEFTKDVGLKEMVPPSSRNPLFTNLDN 1200
1201 LHENNTNQBKKIQERIEKXETLIQENNVLPQIHVTGTGNFMKNLPLSTROKVEGSD 1260
1201 LHENNTNQBKKIQERIEKXETLIQENNVLPQIHVTGTGNFMKNLPLSTROKVEGSD 1260
1261 GAVAPVLODFRSNDSTNRKKTAAHFSKKEEENLEGLGNQKQIVKCYACTTRISPNT 1320
1261 GAVAPVLODFRSNDSTNRKKTAAHFSKKEEENLEGLGNQKQIVKCYACTTRISPNT 1320
1321 SQQNFVQSRKRALQFRLPLEETLEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKE 1380
1321 SQQNFVQSRKRALQFRLPLEETLEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKE 1380
1381 KGATQSPDLSDLTRSHSIPOANRSPPIAKVSGFPSPRIPIYTRVLFDQNSSHLPAASY 1440
1381 KGATQSPDLSDLTRSHSIPOANRSPPIAKVSGFPSPRIPIYTRVLFDQNSSHLPAASY 1440
1441 RKDQSGVQESHFLQGAKKNNLSLAITLWTGQREVGSLGTATNSVTVKQVNTVLP 1500
1441 RKDQSGVQESHFLQGAKKNNLSLAITLWTGQREVGSLGTATNSVTVKQVNTVLP 1500
1501 KPDLPKTSGKVELLPKHVHYQKOLFPPTETNSGSPGHLDLVEGSLLOCTEGAIKWNEANRP 1560
1501 KPDLPKTSGKVELLPKHVHYQKOLFPPTETNSGSPGHLDLVEGSLLOCTEGAIKWNEANRP 1560
1561 GKVPFLRVATBESSAKTSPKLLDPLAWDNHYGTQPKGEWKSQEKSPKTAFFKXDTILSL 1620
1561 GKVPFLRVATBESSAKTSPKLLDPLAWDNHYGTQPKGEWKSQEKSPKTAFFKXDTILSL 1620
1621 NACESNHAIAAINQGNKPEIEVWAKQGRTERLCSONPVLKQHQREITRTTLQSDQEE 1680
1621 NACESNHAIAAINQGNKPEIEVWAKQGRTERLCSONPVLKQHQREITRTTLQSDQEE 1680
1681 IDYDDTTISVMKKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1681 IDYDDTTISVMKKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1741 AQSGSVQFKVWQEBTDSFTQPLRYGELNEHLGLLGPYIRAEVDENIMVTRNQASR 1800
1741 AQSGSVQFKVWQEBTDSFTQPLRYGELNEHLGLLGPYIRAEVDENIMVTRNQASR 1800
1801 PYSFVSSLLISYEEORQAGBPKNFVKNPTKTVFWKQOHMAPTKDEFFDCKAWAYFSDV 1860
1801 PYSFVSSLLISYEEORQAGBPKNFVKNPTKTVFWKQOHMAPTKDEFFDCKAWAYFSDV 1860
1861 DLEKDVHSLGLPILLVCHTNTLPAHGRQVTVQEFALPFTTIDETKSWYFTENMERNCR 1920
1861 DLEKDVHSLGLPILLVCHTNTLPAHGRQVTVQEFALPFTTIDETKSWYFTENMERNCR 1920
1921 PCNTQMEDPTFKENYRFAHNGYIMDPLGLVMAQDQIRWYLLSMGNSNENIHSHFSGH 1980
1921 PCNTQMEDPTFKENYRFAHNGYIMDPLGLVMAQDQIRWYLLSMGNSNENIHSHFSGH 1980

1981 VFTVRKEEYKALYNLYPGVPETVEMLPKAGIWEVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKEEYKALYNLYPGVPETVEMLPKAGIWEVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTEPLGMAHGIRDFQITASQGYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2041 QTEPLGMAHGIRDFQITASQGYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2101 HGKTKGAKQKFSLSLISQFIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSSGIGKNIEN 2160
2101 HGKTKGAKQKFSLSLISQFIIMYSLDGKKQWYRGNSGTGLMVFFGNVDSSGIGKNIEN 2160
2161 PPIIARIYRLHPTHYSIRSLRMLMGCDLNSCMPLMGESKAISSAQITASSYFTNMFA 2220
2161 PPIIARIYRLHPTHYSIRSLRMLMGCDLNSCMPLMGESKAISSAQITASSYFTNMFA 2220
2221 TWSPSKARLELQGRSNAWRPQVNNPKWELQVDQPKTKMKTGVTITQGVKSLTSMYVKEFL 2280
2221 TWSPSKARLELQGRSNAWRPQVNNPKWELQVDQPKTKMKTGVTITQGVKSLTSMYVKEFL 2280
2281 ISSQDQGHQWTLTFQNGKVKVFGQNGDSFTFPVNSLDDPPLITRYLRIHPQSWHQAIRLM 2340
2281 ISSQDQGHQWTLTFQNGKVKVFGQNGDSFTFPVNSLDDPPLITRYLRIHPQSWHQAIRLM 2340
2341 EVLGCBAQDLY 2351
2341 EVLGCBAQDLY 2351

RESULT 2
A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 09-Sep-1994 #sequence revision 09-Sep-1994 #text_change 18-Jun-1999
C:Accession: A47004
R:Elder, B.; Lakiach, D.; Gitschier, J.
Genomics 16, 374-379, 1993
A:Title: Sequence of the murine factor VIII cDNA.
A:Reference number: A47004; MUID:93300511; PMID:8314577
A:Accession: A47004
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2319 <BLD>
A:Cross-references: GR:G05573; NID:G192456; PID:AAA37385.1; PID:G192457
C;Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F;1-19/Domain: signal sequence #status predicted <SIG>
F;23-349/Domain: ferroxidase repeat homology <FO1>
F;402-730/Domain: ferroxidase repeat homology <FO2>
F;1686-2006/Domain: ferroxidase repeat homology <FO3>
F;2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F;2160-2313/Domain: discoidin I amino-terminal homology <DN2>

Query Match 71.1%; Score 8827; DB 2; Length 2319;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;
Qy 1 MQELSTCFELCLLRFCEFSATRYLYGAVELSDWYQSD-IGELPYVDARFPRVPEKSPFP 59
Db 1 MQALPACFFLSFNFCSSAIRRYLYGAVELSNWYIQSDLSLVHLDRLPRKSTSPFP 60
Qy 60 NTSVYVYKTLTFVETDHLFNIAPRPPWMLGPTTQAEVYDVTITLKNWASHPVSJHA 119
Db 61 NTSIMYKTVFVEYKQQLFNIAPRPPWMLGPTTQAEVYDVTITLKNWASHPVSJHA 120
Qy 120 VGVSYWKAASGAYDDQTSQREKDDKVPFGGSHYVQVQLKENGPMASPLCTYSLS 179
Db 121 VGVSYWKAASGDEYDQTSQREKDDKVPFGGSHYVQVQLKENGPMASPLCTYSLS 180
Qy 180 HVDLVKDLNSGLLGALLVCEGSLAKKQTLHKFTLLFAVDEGKSWHSETKNSLMQDR 239
Db 181 HVDLVKDLNSGLLGALLVCEGSLAKKQTLHKFTLLFAVDEGKSWHSETKNSLMQDR 240

240 DAASARAWPMXHTVNGVYNRSLPGLIGCHRSKSVYVHVGMTTPEVHSIFLEHGHTELVRN 299
241 DSASARAWPMXHTVNGVYNRSLPGLIGCHRSKVYVHVGXGTTTPEHSIFLEHGHTFFVRN 300
300 HQASLEISPIITFITAQTLLMDLQGLLSCHSHQHDGVEAYVYKVDSCPEEPQLRMK-N 358
301 HQASLEISPIITFITAQTLLMDLQGLLFCISSHKGHDGMEAYVYKVDSCPEESQWKQKN 360
359 NEZAEEDVDDLTDSEMDVVRDDONSFSFIQIRSVAKKHGPKTWVHYIAAEEDWDYAPLV 418
361 NEEEDYDDDDJY-SEMDFTLLDYDSSP-FIQIRSVAKKYPKTWVHYISAEEDWDYAPSV 418
419 LAPDRSYKSOYLNNGPORI GRKYKVRFPMAVTDDETFRTAIQHSGLIPLLYGEVGD 478
419 PTSDNGSVKSOYLSNGPHRIGRKYKVRFIATYDDETFKTRETIQHSGLLGPLYGEVGD 478
479 TLLIIFKQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFPLPGBEIFKYKWTVTVED 538
479 TLLIIFKQASRPYNIYPHGITDVSPLHARRLPRGIKVLPHPGEIFKYKWTVTVED 538
539 GPTKSDRCLTRYSSFFVNMERDLASGLIGPLLICYKESVDORGNCIMSDKENVILFSVF 598
539 GPTKSDRCLTRYSSFFINPBRDLASGLIGPLLICYKESVDORGNCIMSDKENVILFSIF 598
599 DENRSWLTENIQRP-LPNPAGVQLDEDPFOASNIMHSINGVFDLSQVSLVCLHEVAYWYI 658
599 DENOSWYITENQRP-LPNAKTQPDQPGFOASNIMHSINGVFDLSLELTVCLEHAYWYHI 658
659 LSIQAOTPLSVFSGYTFKHGMVYEDTLTLPFSGETVFMENPGLWILGCHNSDPRN 716
659 LSVGAQTDPLSIFPSGYTFKHGMVYEDTLTLPFSGETVFMENPGLWILGCHNSDPRK 718
719 RGMPTALLKVSSCDKNTGYDEDSYEDISAYLLSKNATPEPSFSONSRHPSTROKOFNAT 778
719 RGMPTALLKVSSCDKNTSYDEIYEDITQVLWENNVDPSFFONTNHPNTRKKFKDS 778
779 TIPENDIBKTDPFWAHRTMPKIQNVSSDDLMLLRQS-PPTHGLSLDLQAKYETETSD 837
779 TIPKNDMEKIBPQEEIAEMLKVQSVSDMLMLLGQSHPTPHGLFLSDGQEAIVEAIIHD 838
838 DPSPCAIDSNLSLSEMTFRPOLHSGGDVFTPEGLQRLNEKLGTTAAATELKLDFKV 897
839 DHPNDAIDSNEGPSKVTQLRPESHSEKIVFTPQGLQLRNKSLETTIEVKWKKLGLQV 898
898 SSTSNMLI-STIPSNLAAGTNTSSLPSPMPVHYDSQDITLFGKSSSPSTESGGPLS 956
899 SSLPSNLMTILSDNLKATPEKTDSSGFPDMPVHSSSKLSITAEKKAYSIVGSHVPLN 958
957 LSEENNDKLLJESGLMNSQESGWKNVSGSTESGRIPKGRAGPALLTKDNALFKVSTSL 1016
959 ASEENSDSDNILDSTIMYSQESIPRDNILSIENDRLIREKRPFGIALLTKDNTLFRDNVSL 1018
1017 LKTKNTSNNSATNRKTHIDGSLILNENSPSVQWNL-ILESDTEFKVTPLIHDRMLMDKNA 1075
1019 MKTKNTYHSTTNEKLHTESP-STENSTTDLOAILKVNSEIQEVTAIIHGDTLILGKNS 1077
1076 TALRLNHNKNTTSKNMNMVQCKEGPTPPDAQNPDMSEFFKMLFLPSPSARMIQRTHGKN 1135
1078 TYVRLNHLNRITSTXKNDIFHRKODDPIPOBENTIMPFSKMLFLSSSNWFKTNGNN 1137
1136 SLNSQGGSPKQVLSLGPESKVEGQNFLEKNKVVGKGEFTKVQVLEKEMVFPSSRNJPL 1195
1138 SLNSQEHSPKQVILMPFKYVKNQSFLEKKNKVTVQDGF-TKNIGLKDMAPFHNKMSIFL 1197
1196 TNLNHLNHNTHNENKIOEBIEKXETLLQENNVLPQHTVGTGNFKNLFLSTRTQNY 1255
1198 TPLSNVHEGNRENGENKIOEBIE-KEALLEEKVLPQVHEA-GSKNPLKDIILILGTRQNI 1256
1256 EGSYDCAVAPVLQDFRSLNDSNTRKKTATHFSEK-KGREENLEGIGNOTKIOVKYACT 1313
1257 --SLVEVHVPLQNITSINNSINTVQIHMEHFHKRKKKETSSEGLVKNKTREWVNY--- 1311
1314 TRISPNTSQCNFVOTORSKALLQKFR-LPLEBETLEKRIIIVDDTSTQWSKNKXHLTSTLTQ 1373

RESULT 3

1312 -----PSQNIITQTSKRALGOFRL-----STQWLKTNCSQTCIIKQ 1349
1374 IDYNEKKGAIITOSPLSDCLTRSHSIPQANRSLPPIAKVSSPFSIRPIYITRVLFDQNSS 1433
1350 IDHSKMKKEITKSSILDS-SVIKSTTQNTSSDSHIVKTSAPP---PIDLKRFPFQNKFS 1405
1434 HLPAAVS---RKDSGVQESSFLQAGKKNLSLAILTLEMTGDQREVSGLSGFSANSV 1489
1406 HVQASSYIYDFKYSSRIQESNPLKETKINNTSLAILPNNMPTDQKFTSPGKSNNSV 1465
1490 TYKVENTVLPKPLPKTCKVELLPKVHIYQKDLFTETSNQSPGHLDLVEGSLLOCTE 1549
1466 TYKKRENIIFLKPTLPESGKIBLLPQVSIQEBEILPTETSHGSPGHLNLMKVFLOKIQ 1525
1550 GAIKMNENANRPGVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEEWKSQSKSPEKT 1609
1526 GPTKWNKAKSHGB--S-LKGTESSKNTRSKLNHAWDYHYAAQIPKDMWKSKEKSEI 1583
1610 AFKKKDTILSNACESNHAIAINEGONKPEIETVNAQGRTERLCSQNPVPLKXHQREI 1669
1584 SIKOEDTILSRPHGNSHSIGA-NEKQWNPQRETTWVKQGTQRTCSQIPVLKXHQREL 1642
1670 TRITLQSDQBEIDYDDTISVEMKKEDFDIYDEDNQSPRFQKTRHYFIAAVERLWDYG 1729
1643 --SAFQSEQBATYDDAITIE-TIEDEFDIYSEDIKQSPRFQKTRHYFIAAVERLWDYG 1699
1730 MSSSPHVLNRAQSGSVPOPKKVVOBFTQSGFTQPIYRGSLNEHLGLGPIYIRAEVEDN 1789
1700 MST-HVLRNRYQDNVFPQKKVVOBFTQSGFQPIYRGSLNEHLGLGPIYIRAEVEDN 1758
1790 IMVTFRQASRPYFYFASLISYEDDQCGAEPKRNFKVKNETKTYFWKVQHIMAPTKDEF 1849
1759 IMVTFKQASRPYFYFASLISYKEDQ-GEZPRRNFKVKNETKTYFWKVQHIMAPTKDEF 1817
1850 DCKAWAFSDVDLEKDVHSGLIGPLLCHVTNTPAGRQVTVQEPALFPIFIDETKSWY 1909
1818 DCKAWAFSDVDLEKDVHSGLIGPLLCHVTNTPAGRQVTVQEPALFPIFIDETKSWY 1877
1910 FTEWERNCRAPCMIOEMDPTFEKNYRFAHNGYIMDTPLGLVMAQDQIRIYWLLSMGNS 1969
1878 FTEWERNCRKTPCMQEMDPTFEKNYRFAHNGYIMDTPLGLVMAQDQIRIYWLLSMGNS 1937
1970 ENIHSIFPSGHVFTVRKKEEYKMAVYNYLPCVFTVEMLPKAGIWRVVECLIGEHLHAGM 2029
1938 ENIHSIFPSGHVFTVRKKEEYKMAVYNYLPCVFTVEMLPKAGIWRVVECLIGEHLHAGM 1997
2030 STLFLVSNKQOTPLGHASGHIRDPOITASQYQGWAPKLABLHYSGSINAWSTKEPESW 2089
1998 STLFLVSNKQOTPLGHASGHIRDPOITASQYQGWAPKLABLHYSGSINAWSTKEPESW 2057
2090 IKVDLLAPMIITHGKTQAGOKFSSLYISOFTIMVSLDGKKWQTVYRGNTGTLMVFFGNV 2149
2058 IKVDLLAPMIITHGKTQAGOKFSSLYISOFTIMVSLDGKKWQTVYRGNTGTLMVFFGNV 2117
2150 DSSGKHNIFNPPIIARYIRLHPHTYSIRSTRMELMGCDLNSCMPJGMBKSAISDAQI 2209
2118 DSSGKHNIFNPPIIARYIRLHPHTYSIRSTRMELMGCDLNSCMPJGMBKSAISDAQI 2177
2210 TASSYFTNMFATWSPSKARLHQBSNANWRPOVNNPKENLOVDFQKTKMKTGVTTQGVKS 2269
2178 TASSYFTNMFATWSPSKARLHQBSNANWRPOVNNPKENLOVDFQKTKMKTGVTTQGVKS 2237
2270 LLTSYVKEFLISSSQDGHQWTLFFQKGVKVFQGNQDSFTPVVNSLDPPLLTLYRLIHP 2329
2238 LFTSMVKEFLISSSQDGHQWTLFFQKGVKVFQGNQDSFTPVVNSLDPPLLTLYRLIHP 2297
2330 QSWVHQIALRMEVLGCEAQDLY 2351
2298 QIWEHQIALRMEVLGCEAQDLY 2319

T42763
coagulation factor VIII precursor - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
R;Lollar, P.
submitted to the EMBL Data Library, August 1996
A;Reference number: 222269
A;Accession: T42763
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2133 <LOL>
A;Cross-references: EMBL:U49517; NID:G1511633; PID:G1511634; PIDN:AAB06703.1
C;Superfamily: coagulation factor VIII; disocoidin I amino-terminal homology; ferroxidase
C;Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-2133/Product: coagulation factor VIII #status predicted <MA>
F;23-345/Domain: ferroxidase repeat homology <FOX1>
F;402-730/Domain: ferroxidase repeat homology <FOX2>
F;1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query March 65.0%; Score 8068; DB 2; Length 2133;
Best Local Similarity 67.0%; Pred. No. 0;
Matches 157; Conservative 208; Mismatches 344; Indels 226; Gaps 17;

Qy 1 MQIELSTCFELCLLRFCSATRRYYLGAVELSWDMOSD-LGELPVDARPPRVKSPFP 59
Db 1 MQIELSTCFELCLLPLGFSAIRYYLGAVELSWYRQSELREHVDTRFPATAPGALPL 60

Qy 60 NTSVYKKTFLVEFTDLHFNIAKPRPMWGLGPTIQAEVYDTVTILKNMASHPVSIHA 119
Db 61 GPSVLVYKKTIVFVEFTDQLFSVARPPPMWGLGPTIQAEVYDTVTVTILKNMASHPVSIHA 120

Qy 120 VGVSYWKASGAEYDDQTSQREKDDKVPFGSHTYVQVLKENGPMASDPLCTYSYLS 179
Db 121 VGVSWKASGAEYDDQTSQREKDDKVLPGKSTYVQVLKENGPMASDPLCTYSYLS 180

Qy 180 HVDLVLDKLSGLICALLVCREGSLAKETQTLHFLFAVDEGKSWHSTKSLMQDR 239
Db 181 HVDLVLDKLSGLICALLVCREGSLTRERTQNLHFEVLLFAVDEGKSWHNSANDSWTRAM 240

Qy 240 DAASARAWPKHVTNGYVNSLPLGLICHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRN 299
Db 241 DPAPARAQAPAMHTVNGYVNSLPLGLICHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRH 300

Qy 300 HQASLSISPTTFLTAQTLMDLQGLSLCHISSHQDGMBAVYKVDSCBEPOLRMKN 359
Db 301 HQASLSISPTTFLTAQTLMDLQGLSLCHISSHHGGMBAHVRVSCAEPQLRFXAD 360

Qy 360 BEABDYDDDLTDSMDVVRFDNDSFPIQIRSVAKXHPKTWHYIAAEEDMDYAPVL 419
Db 361 BE-EDYDDNLYDSMDVVRFDGDDVSFPIQIRSVAKXHPKTWHYISAEEDMDYAPVP 419

Qy 420 APDRSKYSQVLYNNGPQIRGKYKVRPMAYDTFTKTREAIQHSGLGLPLLYGVEGDT 479
Db 420 SPDSRSKYSGLYNSGPQIRGKYKARFVAYDTFTKTREAIQHSGLGLPLLYGVEGDT 479

Qy 480 LLIIIFKNQASRPYNIYPHGITDVRPLYSRRPLKGVKHLKOPPLPGHFKYKWTVTYEDG 539
Db 480 LLIIIFKNQASRPYNIYPHGITDVRPLYSRRPLKGVKHLKOPPLPGHFKYKWTVTYEDG 539

Qy 540 PTKSDPRCLTRYSGSVFNMERDLASGITGPLLIYKESVDQRGNQIMSKRNWILFVFD 599
Db 540 PTKSDPRCLTRYSGSVINLEKDLASGLIGPLLIYKESVDQRGNQIMSKRNWILFVFD 599

Qy 600 ENRWYITENTQRFELPNAGVQLDEPFAQSNIMHSINGYVDSLQLSVCLHEVAYWYL 659
Db 600 ENQWYLAENTQRFELPNAGVQLDEPFAQSNIMHSINGYVDSLQLSVCLHEVAYWYL 659

Qy 660 SIGACTPLSVFSSCYTEKHMYVEDTLTLFPFSGETVPMSENPGLWILGCHNSDFRN 719
Db 660 SVGACTPLSVFSSCYTEKHMYVEDTLTLFPFSGETVPMSENPGLWILGCHNSDFRN 719

720 GWTALLKVSODKNTGDIYEDSYEDI SAYLLSKNALEPRSFSONSRBPTROKOFNATT 779
Db 720 GWTALLKVSODRDIYDNYEDIFGFLLSGKNVIEPRSFSAQNSRPPSASQKQFQIT 779

780 IPENDIEKTPWFARHTPMPKIQNVSSDILLMLRQSPTHGLSLSDIQEAKYTFSDP 839
Db 780 SPEDDVE-LPQSGERTQALEELSVPDGMILLQNPAPHGSSSSDLOEARN-ADDY 836

784 SPGAIDNNLSLSEMTFRPOLHSGDMVFTPEGLQLRLNEKLTGTTAATELKULDFKVS 899
Db 837 LPGAERNTA?SAAARLRPELHSAERVLTPP-----EK-----ELKLDKMS 882

900 TSNNLIS--TIPSDNLAA?DNTSSIGPSPMPHYVDLSOLDTTLFGKSSP?TBSGGPLSL 957
Db 883 SSDLK?TPIPSDLSAETRHSLGPPHPQVNFRLSGALVILGKSSSHF?IGAVFLGS 942

958 SEENDSKLLSGLMNSQSSWKNVSTESGRLFGKRAHPALLTXDNALFVSI-SLL 1017
Db 943 TEEDH-----ESSLGENVSPVESDGI?FEKERAHGPASLTAKDVLFKVNI-SLV 989

1018 KTKTNNSATNRKTHIDGSPSLIENS?SVWQNLBESDTEPKKVTPLIHRMLMDKNATA 1077
Db 990 KTKARVYLKTRKIHIDDAALLTENRASA-----TFMDKNTTA 1028

1078 LRLNHNKNTSSKNVEMVQKKKGPIPPDAQNDPMDFKMLFLPESARLIQRTGKNSL 1137
Db 1029 SGLNHN-----WKGLGKNPL 1047

1138 NSGQSPKOLVSLGPEKSVGQNFSEKNKVVGGEFTKDVGLKEMVFPSSNLFN 1197
Db 1048 SSERGPSPELLTSSGSKSVKSGSGGGRIRVAVEEELSGK--KEMMLPNSBLTFLN 1104

1198 LDMLENNNTNQBKLOEBIEKKETLIOENVLPOLHTVGTGKNPMKQLFLLSTQNV 1257
Db 1105 SADVQGNDSHQKSEEMERREKLVQKVDLPQVYATGATGKTNFLNIHQSTEPSVEG 1164

1258 SYDGAVALQDPRSLNDS?NRKHTAHFSKXGEBENLEGLNOTQI?VEKYACTTRIS 1317
Db 1165 FDGSHAPVQDRLSNDLSAERATHAHFSAIRERAPLAPGRT----- 1210

1318 PNTSQNFVTQSKRALKQFRLPLEETEELKRIIVDDTSTQWSKNMKHLTPSTLTQIDYN 1377
Db 1211 -GPGPSAVERVKQSLQRLPLEEIKPERGVVNLATSTWS----- 1252

1378 EKSGAITQSLDCLTRSHSI?QANRSP?PIAKVSPFPPIRYLTVLFDQNSHLPA 1437
Db 1253 ----- 1252

1438 ASYRKDSGVQESSHFIQGAKNKLSLAILTLEMTGDQREVSGISATNSVTVYKVENT 1497
Db 1253 -----ESSPILQGAKNKLSLFPILTLEWAGGQGGKISALGKSAAGPLASGKLEKA 1301

1498 VLPKPDLPKTSQVVELLPKVIH?YOKDLFPTETNSGSPGHLDLVEGSLLOSTEGA?KNEA 1557
Db 1302 VLSAGLSEASGAEFLPKVRVEREDLLPKTSNVSCAHGDLGOE?IFLQTRGPVNLKV 1361

1558 NRCKVFPFLVATRESSAKTBSKLLDPLAWNHYQTQIPKEWKSQKSPKTA?KKTOTI 1617
Db 1362 NRPG-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALRTKDI 1398

1618 -LSLNACSNHAI?AINEGQNKPEIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLOS 1676
Db 1399 SLPLDRHESNHSIA?KNEGGAET?CREAAWTQKGGPGLCAPKPPVLRHRHORDSLTFQP 1458

1677 DQBEIDYDDTI?SVEMKKBDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGVSSSPHV 1736
Db 1459 EEDKMDYDDIFSTETKGEDFDIYGEENQDPRSFQKTRHYFIAAVERLWDYGVSSSPRA 1518

1737 LRNRAGSGSVQPKKVFQPTGSP?PLRYGELNHLGLLGPY?RAVEDNIMVTRN 1796
Db 1519 LRNRAGSGSVQPKKVFQPTGSP?PLRYGELNHLGLLGPY?RAVEDNIMVTRN 1578

1797 QASRPYSFSLISYEDDSQGAEPKRN?FKVPNETKTYFWKVQHHMAPTKDEDFCKAWAY 1856

Db 1579 QASRPSYSLISYDDQEQABPRNFVOPNETRYFWKQVHHMAPTDEDFCKWAY 1638
Qy 1857 PSDVDLEKDVHSGILGPIVLVCHTWLNDPAHGRQVTVQSEFALFFTFIDETKSWYETENNER 1916
Db 1639 PSDVDLEKDVHSGILGPIVLVCRANTLNAHGRQVTVQSEFALFFTFIDETKSWYETENVER 1698
Qy 1917 NCRAPCIOMEDPDKENYRPHANGVYMDPLGLVMAQDQRIWYLLSMGSGNENIHSIH 1976
Db 1699 NCRAPCHLQMEDPDKENYRPHANGVYMDPLGLVMAQDQRIWYLLSMGSGNENIHSIH 1758
Qy 1977 PSCHVFTVRKEEYKMAVLYPGVFETVEMLPSPKAGIWRVECLIGBHLHAGMTLFLVY 2036
Db 1759 PSCHVFSVRKEEYKMAVLYPGVFETVEMLPSPKAGIWRVECLIGBHLHAGMTLFLVY 1818
Qy 2037 SNKQCTPLGMAHGHRDPTQITASQYQGWAPKLAHLHYSGSINAWSKBPSPWIKVDLLA 2056
Db 1819 SKKQAPLGMASGRIRDFQITASQYQGWAPKLAHLHYSGSINAWSKDPSPWIKVDLLA 1878
Qy 2097 PMITHGKTQCARCKFSLSYSQITIMYSLGKKWQTYRGNSTGTLVFGNVDSSGFKH 2156
Db 1879 PMITHGINTQCARCKFSLSYSQITIMYSLGDRNWSYRGNSTGTLVFGNVDSSGFKH 1938
Qy 2157 NIFNPPPIARYIRLPHTHYSIRSLRMLGCDLNSCMPLGMSKASIDSAQITASSYFT 2216
Db 1939 NIFNPPPIARYIRLPHTHYSIRSLRMLGCDLNSCMPLGMSKASIDSAQITASSHLS 1998
Qy 2217 NMFATWSPSKARLHLQGSNARVQNNPKWQLVDFOKIMKVTVGTTQVKSLTSMYV 2276
Db 1999 NIFATWSPSKARLHLQGSNARVQNNPKWQLVDFOKIMKVTVGTTQVKSLTSMYV 2058
Qy 2277 KEFLVSSQDQGWTLRFQNGKVKVFCQNGQSFVNVKSLDPPILRLYRHPQSWHQI 2336
Db 2059 KEFLVSSQDQGWTLRFQNGKVKVFCQNGQSFVNVKSLDPPILRLYRHPQSWHQI 2118
Qy 2337 ALRNEVLGCEAQDLY 2351
Db 2119 ALRNEVLGCEAQDLY 2133
RESULT 4
KFHUS
coagulation factor V precursor [validated] - human
N:Alternate names: coagulation labile factor; proaccelerin
C:Species: Homo sapiens (mar)
C:Date: 19-May-1999 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000
C:Accession: A56172; A42344; A28028; A27498; A25897
R:Cripe, L.D.; Moore, K.D.; Kane, W.H.
Biochemistry 31, 3777-3785, 1992
A:Title: Structure of the gene for human coagulation factor V.
A:Reference number: A42344; MUID:92232668; PMID:1567832
A:Accession: A56172
A:Molecule type: DNA
A:Residues: 1-2224 <CR>
A:Cross-references: GB:J05368
A:Accession: A42344
A:Molecule type: DNA
2070:2111-2120:2121-2161 <CR>
R:Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A:Title: Complete cDNA and derived amino acid sequence of human factor V.
A:Reference number: A28028; MUID:87260886; PMID:3110773
A:Accession: A28028
A:Molecule type: mRNA
A:Residues: 1-857, R', 859-864, 'R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <EN>
A:Cross-references: GB:M16967
A:Note: parts of this sequence, including the amino end of the mature protein, were dete
R:Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
Biochemistry 26, 6508-6514, 1987
A:Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A:Reference number: A27498; MUID:88107560; PMID:2827731
A:Accession: A27498

A:Molecule type: mRNA
A:Residues: 1-1284, 'I', 1286-1600 <XAN>
A:Cross-references: GB:M17785
R:Kane, W.H.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A:Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A:Reference number: A25897; MUID:86313665; PMID:3092220
A:Accession: A25897
A:Molecule type: mRNA
A:Residues: 1188-1215, 1315-2224 <KA2>
A:Cross-references: GB:M14335
R:Keller, F.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A:Title: Thrombin-catalyzed activation of recombinant human factor V.
A:Reference number: A56139; MUID:95210278; PMID:7696276
A:Contents: annotation; thrombin cleavage sites
C:Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C:Genetics:
A:Gene: GDB:F5
A:Cross-references: GDB:1119896; OMIM:227400
A:Map position: 1q23-1q23
A:Introns: 53/2; 84/1; 125/1; 136/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65
C:Function:
A:Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A:Pathway: blood coagulation
C:Superfamily: coagulation factor V; discoidein I amino-terminal homology; ferroxidase re
C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma; ;
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2224/Product: coagulation factor V #status predicted <MA7>
F:29-737/Product: coagulation factor Va heavy chain #status experimental <VAH>
F:29-345/Domain: A1 <DA1>
F:33-329/Domain: ferroxidase repeat homology <PO1>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <PO2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-I-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1561-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>
F:1906-2061/Domain: discoidein I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2221/Domain: discoidein I amino-terminal homology <DN2>
F:51-55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,149
F:167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/Disulfide bonds: #status
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363,693,1546/Binding site: sulfate (Tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382,1338/Binding site: carboxylate (Asn) (covalent) #status absent
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 22.7%; Score 2823.5; DB 1; Length 2224;
Best Local Similarity 30.6%; Pred. No. 2.2e-139;
Matches 769; Conservative 388; Mismatches 81; Indels 507; Gaps 77;
Qy 22 RRYILGAVELSWDYMOSDGLGELPVDARFPPRPVPSFPNTSVV-YKKTILFVEFTDHLFNI 80
Db 32 RQFYVAAQGISWSYRPE-----PTNSSILNLSVTGFKKIVREPY-FKK 75
Qy 81 AKRPPPMGLLGPFIQAEVDVTVVITLKNMASHVPLHANGVSTWAKSEGEYDDQSQR 140
Db 76 EKPOSTISGLGLTYAEVGGIIRHFKNKADKPLSHPOGIRYVKSLSEGSYLDTHTPPA 135
Qy 141 BKEDDKYFPGSGHYVWQVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCRE 200
Db 136 ERMDDAVAPGREYTYEWSISDSGFTDDPCLPLHIYYSHENLIEDFNSGLIGPLICKK 195

201 GSLAKEKTO-TLHK-FILLPAFDEGKSWHGETKNSLMQDRDAASARAWPKMHTVNGYN 258
196 GTLTEGGTQKTFDQOIVLLFAVFBESKWSQSSS-----LMYTVNGYN 239
259 RSLPGLGCHRKSVYHVGTTPEVHSIFLEGHTFLVRNHRQASLEISITFLTAQTL 318
240 GTMPDITVCAHDHISWLLGNSGPELFSIHFGQVLEQNHKKVSATLVSATTTANMT 299
319 LMDLQGF-LSCHISSHQHDGMEAVKVYDSCPEEPQLRMKNNEBAEDYDDDLTDSMDVVR 378
300 VGPEGKWIISLTPXHLQAGQOAVIDIKNCCKTRNLKITYRE----- 342
379 PDDNSPFIQIRSVAKKHPTWYHIAAEBEDWDYAPLVLPADDRSYKSOYLANGPQRI 438
343 -----QRHMKRWEVYFAAEVIVDYPVPIANMDKKYRSOHLNFSNQI 387
439 GRKYKVRFMAYTBTETKTRAI---QHESGILGPLYGEVGDILLIFKQOARPNYIY 495
388 GHYKKVMYTOYEDES-FKHTVNPNMKEDGILGPIIRAQVRDTLKIYFKNWSRPYSIY 446
496 PEGIT-----DVRPLYSRRLPKGVKHLKDPILPGETFKYKWTVVEDGPTKDPCLT 549
447 PHGVTFSYEDVAVSSFT-----SGRNTMIRAVQGETIYTKNLELFEDEPTDAQCCT 502
550 RYSSFVNMEZDLASGLIPZLIICYESVDQRGQIMSDKRNVLFSVFDNRKSWILTEN 609
503 RPYSDVDIMEDIASGLIGLILCKSRSLDRGQRAADIEQOAVFAVDENKSWYLEDN 562
610 IQRFLPNAGVLEDPBPAQSNHSHINGYVFDL-QLSVCLHEVAYWYILSIGAOTDPL 668
563 INKCEMFPDEKRDDBKPYESINMSTINGYVPESTITLGFCDFTVQWHFCSVGTQNEIL 622
669 SVFSGYTFKHVYEDTLTPPESGETVFMSPNGLWILGCNSDPNRGMTALLKVS 728
623 TIHTGHSFYGKHEHDTLTPRMRGESVTVMONVGTWMLTSMNSPRSKLRKLRDV 682
729 SCDKNTGDYEDSYEDI---SAYILSX--NNAIEP-----RSF 761
683 KC---IPDDDESYEIPPPSTVWATRKWDRLEPEDESDADYDQNRALAAALGINSF 739
762 SONRHPSTROKORWATIP-ENLIEKTDPPFAHTPMKTONVSSDLMLLQSPFPH 820
740 RNSLI--NQEEENLTALENCTE-----FVSSNTDIIVGNSYSPS 781
821 GLS-----LSDLOSA---KYETPSDDPSGALDSNNLSLSEMT--HPRPOLHSDMVFT 869
782 NISKFTVNNLAEPQKAPHQQAATAGSLRLHIGKSVLNSSTABESP---YSDPTIED 838
870 P-----ESGLQRLNEKLGTTAATELKLJFKVSVSTSNLITIPSDNLA----- 914
839 PLQPDVTGIRL-----LSLGAGEFKSQEHAHKGPVERDQAAKHFPMWKLIA 887
915 --AGTDNTSSCGPPS-MPVHYDSQDITLFGKSSPLTESGPLSLSENDSKLL----- 967
888 HKVGHLSQDTGSPSGMRPEWDLPSQDTGSPSRMPKWDPPSDULLLKQSNSSKLLVGRW 947
968 ---ESG-----LMNSQESS--WKG-----NVSTSEGRALFKGKRAHG 999
948 HLAESKGSYEIIQDDEDTAVNNWLISQNASRANGESTPLANKGQSGHPKPRVHK 1007
1000 PALLTKO--NALFKVSIISLLTKNTKSNASNRKTHIDGPSLLIENSPSVQNLIESDTE 1057
1008 SLOVRQDQSGRLKSKSQFLIKTRKKKXKHTH-----APLSRTEFHLRSEAYN 1057
1058 FKVVTPLAHDRLMIDKNATALRHNMSKNTSSKNMVMQKKGCFIPDQCN--PDMSF 1115
1056 TFSEERLKHSLVL-----HKSNETS-----LPTDLNQLTSPSMDP 1091
1116 FKMFLPESARMIQTHGKNSLN-SGQGFSPKQVSLGPEKSVEQCNF-LSEKKNVWVGK 1173
1092 GWIASLPD-----HNCNSNDTGOACSPGLYCTVPEE-HYQTFPIQDPQX---- 1138
1174 GEFTKDVG-----LKEMV-FPSSRNFLTLNLDNLHNHTNQEKKI-----QEEIEKKE 1221

1139 -HSTDSHRS SSPELSEMLEYDRSHKSFPTDISQSPSSSEHEVWQVISPDLQVITLSP 1197
1222 TLQENVVLQIHTVGTQKRFMKNLFLLLSTRQNVESYDQAYAPVLQDFRLNDSTNRT- 1280
1198 ELSQTNLSPDLSHTTSLFELIQRNL-----SPALGOMFISPLDLSHTL 1240
1281 ---KHTAHFSKKEEENLEGLNQTKQIVKVIACCTTRISNTSQCNFVTQSRKALKQF 1337
1241 SPDLSHTT-----LSLDLQ-----TNLSPELSQTNL-----SPALQO- 1273
1338 RLPLEETELEKRIIVDDTSTQWSNMKHLTPS-TLTQIDYNEKXKGAITQSPSLDCL--- 1393
1274 -MPLSPDLSHTTSLDPSQTNLSPELSHMTLSPELSQTNLSP-----ALGOMFISPLDLSHT 1328
1394 TRHSIQANRSPPIAKVSPSPSIRPIYILTRVLFDQNSSHLPAASYRKKDQSGVQBSHP 1453
1329 TSLDPSQTNLSP-ELSQTNLSPALGOMPLS-----PDPSTTSLDLSQT--- 1373
1454 LQAKKNLSLAILTLEMTGQREVGLSGTSATNSVTYKVKVENTVLP---KPDLPKTSKG 1510
1374 -----NLSPELSQTNLSPDLSMPPLFADLSQIPLPDLDQMTLSPDLDGTDLSNPGQ 1426
1511 VELLPKV-----HIYKDLFPPTETNSGSPGHLDLV-----EGSLQTEGAIKWNE 1556
1427 MSLSPDLSQVTLSPDLSQTNLSPDLSQTNLSPDLSQTNLSPDLSQTNLSPDLSQTNLSP 1480
1557 ANRPGKVPFLRVATESSAKTE-----SKLLDPLAW-----DNHYGTOIPKEKWSOB 1603
1481 S-----FPYDLGOMPSPSPTLNDTLSEFNFPLVIGLKGDTYIIBILPREEVOSES 1535
1604 KSPKTAFFKKDITLSLNACESNHAIAINEGQNKPEIETWAKQGRTERLCSQNPVYLK 1663
1536 ----- 1535
1664 RHQREITRTLQSQOBEID---YDDTISVEMKKEDFDIYDEBENQS---PRSFQKKTTHYF 1718
1536 -----DYAEIDYVPYDDPYKTDVTRINSSRDPDNIAAWYLRNNGNRNRY 1583
1719 TAAVERLWDYCMSSSPHVLNR--AQSGSVQ---FKKVQFQETDGSFTQPLVREGELNE 1773
1584 TAAEISWDY---SEFVQRETDIEDSDIDEDTYYKVVFPKRYLDDSTFKRDRGEYEE 1639
1774 HGLLGLYVIRAEVDNIMVTPRNOASRPFYSYSLISYE-----EDRQGAEPKKNF 1825
1640 HGLILGPILRAEVDVQVRFKNLASRPSYLAHGLSVKESSEKGTVEEDDSPEWKEDNA 1699
1826 VKPNETKTYFKVQHHVAPKDFEDCKAWAYFSDVLEKDVHSGLIGLPLLVCHTNTLNP 1885
1700 VQPNSSYTYVWHATERSGSPSGSACRAWAYSVANPEKDIHSGLIGLPLLVCHTNTLNP 1759
1886 HGRCVTQEPALFTTIFDETYSVFTENMERNCRAPCNIOMEDPTFENYFHAINGYIM 1945
1760 SNMFLMREFFVLLPMTFDEKKSUYEKKSSRW-----LTSSEMKSHFHAINGYI 1813
1946 DTLFGLVMAOQRIWYLLSMGNSNIHSHFSGHVFTVRKERYKVALNLYPGVFTV 2005
1814 -SLPGLXWYEQEWRVLLHLLNIGGSODIHVHFHQTLLENGKQHLQWGLPPLGSPKTL 1872
2006 BMLFSKAGIWEVECLGEHLHAGHSTLFLVYSNKQOTPLGMAHSHIRDFOITASQVQW 2065
1873 BMKASKPGWMLLNTVEGENQAGQTPFLIMDRCRMPLGSLTGIIISDSQIKASEFLGYW 1932
2066 APKLARLHYSSINAWSTKE---PFS---WTKVLLAPMIHGIKTQAROKFSSLYISQ 2119
1933 BPRLARLNGSGYNANSVEKLAEPFASKPILQVDMQKEVILITQTOGAKHYLSCTYTE 1982
2120 FIIMYSLDGKQWYRGNTGTMLVFFGNVDSSGKHNIENPPIIARYIRLHPHYTHSIRS 2179
1993 FYVAYSSNQINQWQFKGNSTENVMYFNGNSDASTIKENQFDPPIVARIISPRAYNR 2052
2180 TLRBELGCDLNSCGMPLGMESKALISAOITASSYFTNMFAT-WSPSKARLHLGORSNAW 2238

Db 2053 TLRLLEQCEVNGCSTPLGMEKNGKIENKQITASSPFGKSGWGDYWEPPARLNAAQGRVNAW 2112
 Qy 2239 RPQVNNKEMLOVDFQKMTKVTGTTQGVKSLLSMYKEFLISSQDCHQWTHFFQNGK 2298
 Db 2113 QAKANNKQWLEIDLLKFKITAITQCKSLSEMYKVSITIHYSEQGVWKPRLKSS 2172
 Qy 2239 V--KVFGKQDSFPVNVNSLDPPLTLRLIHPQSVWHQIALRMEVLGCEAQDLY 2351
 Db 2173 MVDKIFEGNTYKGVKVNFPFPIISRIVRIPKTNQNSIALRLFLFC---DIY 2224
 RESULT 5
 T42764
 C: coagulation factor V - mouse
 C: Species: Mus musculus (house mouse)
 C: Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
 C: Accession: T42764
 R: Yang, T.L.; Cui, J.; Rehmtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg, Blood 91, 4593-4599, 1998
 A: Title: The structure and function of murine factor V and its inactivation by protein C
 A: Reference number: 222270; MUID: 98282202; PMID: 9616155
 A: Accession: T42764
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: mRNA
 A: Residues: 1-2183 <YAN>
 A: Cross-references: EMBL:U52925; NID:G3219690; PID:G3219691; PIDN:AAC99553.1
 C: Function:
 A: Pathway: blood coagulation
 C: Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C: Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma; F350-652/Domain: ferroxidase repeat homology <FOX1>
 F: 1541-1864/Domain: ferroxidase repeat homology <FOX2>
 Query Match 22.3%; Score 2774.5; DB 2; Length 2183;
 Best Local Similarity 29.9%; Pred. No. 8,1e-137;
 Matches 750; Conservative 408; Mismatches 851; Indels 503; Gaps 73;
 Qy 8 CPFLCLL---RFC-----PSATPRVYLGAVELSDYMQSDIGELPVDARFPRPEKS 56
 Db 7 CFFLLVLTGRWAGSGHQAQAQLRQFVAAQGLVNVHPE-----PTD 51
 Qy 57 PPFNTSVYVYKTLFVETDHLFNIAKPRPPMGLIGPTQIAEYVTVVITLKNASHVVS 116
 Db 52 PSLNIPSPFKIVVREY-EQYFKKPRSSNGLLGPTLYAEVGDVIVKHFENKADKELS 110
 Qy 117 LHAVGVYWKAGEAGYDDQTSCKEKEDDKVPGGSHTVVMOVLKENGEMASDPLCLTVS 176
 Db 111 HPQGIKYSKSEGAASYADHFPAPKDDAVAPGEYIYEWIVSDDSGTDDPPCLTHI 170
 Qy 177 YLSHVELVKDLSGLIGALLVCREGLAKKXTQL--HKPILLFAVDFEGKSWHSETKNS 234
 Db 171 VYSYENLTQDFNSGLIGPLLI CKKGLTBDGTQKMPDKQHVLLFAVDESKS----- 222
 Qy 235 LMQDRDASARAPKMTHTVNGVYVNSLGLICHKRSVYVHVGITPPEVHSIFLGHT 294
 Db 223 -----RSQSPSLMYTINGVNTMPDITVCAHDHVSWHLIGMSGELFSIHFNQGV 274
 Qy 295 FLVNRHQASLISPIITFLTAQTLMDLQCFLLSCHISHQDGMAYVKVDSCEEPOL 354
 Db 275 LEQNHKVSVTVLVGSATSTANWTSPEGRWIVSSLI PKHYQAGQCAVIDKNCEPKTR- 333
 Qy 355 RMKNNEAEYDDDLTDSMOWVREDDNSPFIQIRSVAKKPKTWVHYIAAEEDWDY 414
 Db 334 -----SPK--TLTREQRMYKREWEYFIAAEVWNY 362
 Qy 415 APLVLAPDDBRSKSOYLNNGPQIRGKYKKRFMAYTDETF--KTREAIQHSGLGPLL 472
 Db 363 APVIPAAMDKIYSEQLDNFSNQIGKHYKYVYRQYEEETFKRTDNPISIKOSGLGPVI 422
 Qy 473 YGEVGTLLIIFKNQASRPNTYPHGITDVRPLYSRRLPKGV-----KHLKDRPILPG 525
 Db 423 RAQVRDTLKVFNKVASRRYSIYPHGV-----RSPYEDGINSSTSGSHITIRVPQG 476

Qy 526 EIFKYKVTWVEDGPTKSDPRCLTRYYSFVNMERDLASGLIGPLLI CYKESVDQRNQI 585
 Db 477 ETFFYKNILFEDEPTENDAQCULTRPIYSDVVDTRDIASGLIGLLI CKSRSLDQGVQR 536
 Qy 586 MSDKRNVLFSVDENRSWYLTENIQIFLNPAGVQLEDPEFQASNMIMISINGVYFSLQ 645
 Db 537 VADIEQQAFAVFDENKSVIENKFCENPDVDKDDPKFYESNTMSTINGVVPSSIS 596
 Qy 646 -LSVCLHEVAVWYILSJCACTDFLSVFPSTGVTFKHKVYEDTLTLFSPSGTVMSENP 704
 Db 597 TLGFCFDOTVQWHEFCSVTHDDILTIHFTGSHFYIGRRHEDTLFLFPMRGESVTVMNV 656
 Qy 705 GLWILGCHSNFRNRGMTALLKVSSCDKNTGDY-YEDSYE----- 743
 Db 657 GTWMLTTMNSPKERNLRLRFDDVKNR---DYDNEDSYBIYEPAPTSMTTRIHDSLE 713
 Qy 744 -----DISAVLLSKNNAIEPRFSQNSRRHSTKQKFNATIP-ENDIEKTDWFA 793
 Db 714 NEFGIDNEDDDYQYLLASSLGI--RSPKNSSLNP--EENEFNLTALENSSEFISP--- 766
 Qy 794 HRTMPKIQNVSSDMLMLLRQSTPHGLSLSD-QEAKYETFSDDPSFGA----- 843
 Db 767 ---STDRVVDNSRILSKIINN-----NLKQFQTL-----PGCATVAGTLLRNL 810
 Qy 844 --IDSNNLSLSEMTFRPQLHSGDMVFTPEGLQRLNEKLGTTAATLKKLDKFKVSSTS 901
 Db 811 IGLDENFVLNSTRHRSSSYHENDME-NPQSNITWVYLLPLGPKG----- 854
 Qy 902 NNLSTTPSDNLAAGTNTSSLGPPSPMPVHVDQIDTTL--FGKSSPLTBSGGPLSLSE 959
 Db 855 -----SGNREQDKPKTIKTGRPHMKHFRFWMKAPAGKTGRHSPNPKNSYSG--MKSE 904
 Qy 960 ENNDSKL-----LESGLMN-----SQESSW---GKNVSTESGRLFKGRAGPALLT 1004
 Db 905 EDISELPLZOKTSTFELNRWRVASEKSYELIANGCEDTVDXLI----- 951
 Qy 1005 KDNALFKVSIILLKNTKNTSNNSATNRKTHIDGPSLLIENSFVWQNLIESTEFKKVTPL 1064
 Db 952 -TNSPQONITVPGESTSHNTTKPSDLPFTPSGVGHKS PHVQEQ--EENSQGKQGLF 1008
 Qy 1065 IHRMLDKNATLRLNEMSNKTTSSKNMEXVQCKEGPIPPDAQNPDMSPFKMLFLPES 1124
 Db 1009 IRTI-----KKKNKKLAL-----HSPLSRPGFDP----- 1033
 Qy 1125 ARWIQTHGKNSLNSGGQSPKQLVSLGPESXVEGQNFLEKKNVVGKGEFTKDVGLKE 1184
 Db 1034 -----LEGHNHP----- 1041
 Qy 1185 MVFPSSRLFTNLNDLHNHNTNQEKIQEIEKKETLIOENVVLPQI--HTVTGTKNF 1242
 Db 1042 --FPDRLLNHSLL--LHKSN---ETALSPLANQTSPEMSTDRSLPDYNOYSKNDTSQM 1093
 Qy 1243 MKNPLFLSTRQNVGSDYGAAPVLQDFRSLNDSNTNRKHTAHFSKKEEENL----- 1296
 Db 1094 SSSLDLY---QSVFAEHSPTFPA-QDDQTHSTIDPSYRSSPPELSQGLDYLDLSDHFP 1149
 Qy 1297 EBLGNQTKQIVKQYACTTRISPTNSQQNFVTRSKRALXQFRLPL--BETLEKRII--- 1351
 Db 1150 DDIG-----LTSFPFDQSKSSFSDDQAI PPSDLSLFTTISPDLQTIIPD 1197
 Qy 1352 VDDTSTQWKNKXHLTPSTLTQI-----DYNBK-----EKGAITQSPSLDCLTRSHIPOAN 1403
 Db 1198 LDQLLSPEDKQKTSPP-LGQVFLSPDDNQKTSPPDLGOVLSLPPDDNQKTS---PDLG 1253
 Qy 1404 RSLPLI---AKVSPSPSIRPIYLT-----RVLFQDNSSHLPAAVYRKDKSGVQESSHFLQ 1456
 Db 1254 QVFLSLDDNQKTS--PDLGQVFLSPDDNQMTSPDLGQVPLSSNQKTS SPDLQGVFLFP 1312
 Qy 1457 AKKNLSLAIITLMTGQREVSGSLGTSATNSVTYKVENTVLPKPD-----LPKTSKV 1511
 Db 1313 EDNQVFLDLQVPLSSDQO---STSTDLTLSPDFGQTVLSPDLQQLPFDNSQV 1368
 Qy 1512 ELLPKVH-----IYQKDLFPPTETNSGSPCHLDLVEGSLLOQTEGAIKWNEANRP 1560

```

Db      1369 TVSPDLSLLTSPDFNEIILAPDLGQVTLSP-----DLQTN----- 1405
QY      1561 GKYPFLRVATESSAKTSPKLLDPLANDNHYGTQIPKEWKSQKSPKTKFAKXDTILSL 1620
Db      1406 ---PALNHGHCASSADPDQASYSP--DSGQASSLP--ELAKTILPHPLDTHIPPPSPSPTL 1458
QY      1621 NACESNHAIAINEGQNKPIEIVTWAKOGTERLCSQNPV---LKRHQREIRFTVLOS 1677
Db      1459 N-----NTLSRKFNPLVWVGLSRVGDGVVEIPSE 1490
QY      1678 QBEIDYDTISVEMKK3DFDIYDEDENQSPRSFOKKTRH-----YFI 1719
Db      1491 PERLDED-----YAEEDFTYNDPYTDRTDWNSSRPDTTAAWYLRGSGHKPFYI 1544
QY      1720 AAVERLDWYGMSSSPHLNRNAGSGVQP---FKYVQFBFTGSGFTQPIYRGLNHLG 1776
Db      1545 AABBITWNAEFAQSEM--DHEDTGH7PKOTTKVVKVFRKYLDTSTFRDPRAEYEBHLG 1602
QY      1777 LIGPYIRAEVEDNIMVTFRNOASRPSPYSLSIYE-----EDQEQAGABPKFNKVP 1828
Db      1603 ILGPVIRAEVDDVIQVAFKNLASPPYSLHAHGLSYEKSSEKTYEDBSPEWFODDAVQP 1662
QY      1829 NETKTYFWKQHNMAPTKDFCDKAWAYFSDVLEKDVHSGLIGLPLVCHTINTLPAHGR 1888
Db      1663 NSSVYVWHAATKSGPENPGSACRAWAYSAVNVYERDIHSLIGLPLICRKGILMERNL 1722
QY      1889 QVTVOEALFTTIDETKSYWFTENMERNCAPNIOEMEDPTFKENRFRHAINYIMDTL 1948
Db      1723 FMDMRPEVLLFMVDEKSKWYEXS--KGRS-----RIESPEERNAHKFAVAINMIYN-L 1774
QY      1949 PGLVMAQDQIRVLLSGNSNTHSHFSGHVFTVRKREYKMALNLYPGVETVEML 2008
Db      1775 PGLMYDQEWRLHLLNMGSRDIHVHPRGQTLDRTKQHQGLGWPLLPGSKTLEMK 1834
QY      2009 PSKAGIWRVECLIGHLAGMSTLFLVYSNKKCTPLGNVAGSHIRDFQITASQYGQWAPK 2068
Db      1835 ASKPGWLLDTEVGENGQVAGQWQTFPLIDCKCPMGLSTGVISDSQIKASEYLTYWEPR 1894
QY      2069 IARLHYSGSINAWTKE-----PRS-WIKVDLLAPMIHGHKTGQAKQKSSLIISQFII 2122
Db      1895 LARLNNAGSYNAWSTIEKTALDFFPKPIQVDMQREVVVGTQCGAKHYLKSCTFTFQV 1954
QY      2123 MYSLDGKKWYRCNCTLTMLVFGNVDSSGKINFNPRPIAIVIRLHPHYGIRSTLR 2182
Db      1955 AYSDDQTNQWLFGRGSKSVWYFTGNSDGTIKENRLDPPVARIYIRLHPKYSNRPTLR 2014
QY      2183 MELMGCLNSCSMPLGMESKAISDAQITASSYFTNMFAT-WSPSKARLHLQGRNWRPQ 2241
Db      2015 LELQGEVNGCSTPLGLEDRIGQKQITASSFKKSGWGDYWEPSLARLNAQGRVNAWQAK 2074
QY      2242 VNNPKWLVDFQKTMKVTGVTQGVKSLTSMYVKEPLISSQDGHOWTLFFQNGKV-- 2299
Db      2075 ANNNKOMLVDLTKTKKTAIVTQCKSLSEMYKYSIQYSDQGVNAWKPYROKSMVD 2134
QY      2300 KVFQGNQDSFTPVNNSLDPPLTLRYLRHPSQSVWHQIALRMEVLGCEAQDLY 2351
Db      2135 KIFGNSNTKGMKNFPPIISREIRIIPKTNQSIARLELPGC---DIY 2183

```

RESULT 6

```

KPB05
coagulation factor V precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C/Accession: A42580; A36497
J/Guinito, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A/Title: The complete cDNA sequence of bovine coagulation factor V.
A/Reference number: A42580; MUID:92147638; PMID:1737753
A/Accession: A42580
A/Molecule type: mRNA
A/Residues: 1-2211 <GUI>

```

```

A/Cross-references: GB:M81440; NID:gi63037; PIDN:AAA30512.1; PID:gi63038
A/Note: sequence extracted from NCBI backbone (NCBIN:80774, NCBI:P:80776)
R/Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A/Title: Identification and characterization of a phospholipid-binding site of bovine fa
A/Reference number: A36497; MUID:91072354; PMID:2254316
A/Accession: A36497
A/Molecule type: protein
A/Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 1681, 'X', 1
Biochemistry 33, 13109-13116, 1994
A/Title: Determination of the disulfide bridges in factor Va heavy chain.
A/Reference number: A55979; MUID:95034740; PMID:7947716
A/Contents: annotation
A/Note: 566-Cys and 617-Cys were shown to have free sulphydryls
C/Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C/Function:
A/Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot
A/Pathway: blood coagulation
C/Superfamily: coagulation factor V; discoidein I amino-terminal homology; ferroxidase re
C/Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma; ;
F/1-29/Domain: signal sequence #status predicted <SIG>
F/29-2211/Product: coagulation factor V #status predicted <MAT>
F/29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
F/29-345/Domain: A1 <DA1>
F/33-329/Domain: ferroxidase repeat homology <FO1>
F/346-695/Domain: A2 <DA2>
F/351-688/Domain: ferroxidase repeat homology <FO2>
F/656-1564/Domain: B <DOB>
F/1175-1437/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F/1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F/1565-1892/Domain: A3 <DA3>
F/1572-1892/Domain: ferroxidase repeat homology <FO3>
F/1854-1752/Region: phospholipid binding #status predicted
F/1883-2051/Domain: C1 <DC1>
F/1893-2048/Domain: discoidein I amino-terminal homology <DN1>
F/2052-2211/Domain: C2 <DC2>
F/2052-2208/Domain: discoidein I amino-terminal homology <DN2>
F/167-193, 248-329, 499-525/Disulfide bonds: #status experimental
F/225, 329, 397, 382, 460, 553, 587, 745, 756, 774, 780, 902, 952, 964, 1044, 1053, 1062, 1071, 1078, 1094,
F/334-335/Cleavage site: Arg-Asn (protease C) #status predicted
F/363, 657, 1537/Binding site: sulfite (Tyr) #status predicted
F/376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F/533-534/Cleavage site: Arg-Gly (protein C) #status predicted
F/607-688, 1712-1738, 1894-2048, 2053-2208/Disulfide bonds: #status predicted
F/741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F/1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predicted
F/1564-1035/Cleavage site: Arg-Ser (thrombin) #status experimental

```

```

Query Match      22.1%; Score 2747; DB 1; Length 2211;
Best Local Similarity 30.2%; Pred. No. 2,3e-135;
Matches 765; Conservative 382; Mismatches 833; Indels 550; Gaps 80;

```

```

QY      22 RRYLGVAVELSWDMQSDGLPVDARPPRPVPSFPFVSVVYKTLFVFTDHLFNIA 81
Db      32 RQFVAAQSRINWYR-----PESTHL-----SSKPFETS--FKKIVYREY-EAYPKE 76
QY      82 KPRPPWMLGLPTTQAEVYDVTVTLLKNMASHPVSLHAGVGVWYKASGEAYDDQTSORE 141
Db      77 KQSRTEGLGLPTLYAEVGLMKVHFKNKAKPLSIHAQIKYKFKSEGASYSHTLPM 136
QY      142 KEDDKVPFGSHTYVWQVLKENGPMASDPGLCLTYSYLSHVDLVKDLNSGLIGALLVREG 201
Db      137 KMDDAVAPGQPTTVEYIISHSRSGPTDDPPCLTHIYYSVNVLVEDFNSGLIGPLICKG 196
QY      202 SLAEKTKQTL--HKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKHTYGVNR 259
Db      197 TLTDGQKMPKQCHVLMFAVDFDESKSNQTS-----LMYTVNGYNG 240
QY      260 SLPGIGICHRKSVVWHVIGMGTTPPEVHSIFUEGHTFLVRNHRQASLEISPTFTTAQTL 319
Db      241 TMPDITVCAHDEISWHLGMSGGPELFIHFNGQVLEONHHKISAITIVSATSANTMTV 300

```


Db 2145 SDQCTDWKPYREKSMYDKIPEGNNVGHVKNFNPFIIRFIRIIPKTIWNQSIARLRL 2204

Qy 2342 VLGCERADLY 2351

Db 2205 LFQGC---DMY 2211

RESULT 7

A25945

coagulation factor VIII - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000

C:Accession: A25945

R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Mutha, P.; Wasley, L.C.; Kaufman, R.J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986

A:Title: A large region (approx 95 kba) of human factor VIII is dispensable for in vitro

A:Reference number: A25945; MUID:86287369; PMID:3016730

A:Accession: A25945

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-869 <TOO>

C:Superfamily: coagulation factor VIII; disocoidin I amino-terminal homology; ferroxidase

Query Match 17.6%; Score 2190; DB 2; Length 869;

Best Local Similarity 45.2%; Pred. No. 7.8e-107;

Matches 493; Conservative 122; Mismatches 251; Indels 224; Gaps 15;

Qy 705 GLWILGCHNSDFRNGMTALLKVSSCDNKGYYEDSYEDISAVLLSKNAIEPRSFQSN 764

Db 1 GLWILGCHNSDLRNGMTALLKVSSCDNKGYYEDSYEDIPGFLSKNVIEPRSFQSN 60

Qy 765 SRHSTROKQFNATVTPENDIEKTDPPFAHRTMPKIQNVSSDILLMLROSPTPHGLSL 824

Db 61 SRPFSASQKQQTITSPDDVE-LDPQOERTQALELSVPSGDSMLLQNPAPHGSS 119

Qy 825 SDLOAKYETSDPSGAIDNNLSSEMTFRPOLHSGMVFPTPESGLQIRLNEXLGT 884

Db 120 SDLOEARNE--ADDYLPAGARENTPASAAALRPELHHSABRVLTPEP-----EK--- 167

Qy 895 TAAVELKGLDKVSTSTNNLS--TIPSDNLAAGTDNTSSIGPPSPVHYDSOLDTILFG 942

Db 168 ----BLKKLDSKMSDDLKTSPTIPSDTSAERTHTSLGPFPQVNSQLGAIVLG 223

Qy 943 KKSSPLTESGGLSSENNSKILESGLMNSQESWGVSTSGRLFKGKRAHGPA 1002

Db 224 KNSHFIAGVPLGSTEEDH-----ESSLGENVSPVSDGIFEKERAGFAS 270

Qy 1003 LTKNALPKVSIILKTKNTSNSTNATNKTHTDGPSSLIIENSPSVQWQMLIESDTFKKVT 1062

Db 271 LTKDVLFPKVISLVKTKARVYLKTRKIHIDDAALLTENRASA----- 315

Qy 1063 PLIHDRMLNDKATRLRNHNSKNTSSKNMVMQKKEGPTPPDPAQNPDMSFFKMLFLP 1122

Db 316 -----TFMDKNTTASGLNHVSN----- 332

Qy 1123 BSARWIQRTGKNSLNSGGPKQPKQLVSLGPKKVEGGQNFLLSEKKNVYVKGFTKDYGL 1182

Db 333 ---WKGLPKGNPLSSREGPSPPELLTSSGSKSVKVGSGQGRIRVAVEBELSKG--- 385

Qy 1183 KMWVPSRRNLPLTNDLNHNNTINOEKIQEEIEKKEETLIQENNVLPQIHVTGTGNF 1242

Db 386 KEMLPNSBELTFLTNSADVQGNTHTSQGGKSEEMERREKLQVKVLDLPQVVTATGTKNF 445

Qy 1243 MKNLFLTSFRQVVEGSDGAYAPVLQDFPSINDSTNRTKHTAHPSKKEEENLEGLNQ 1302

Db 446 LRNIHQSTEPSVEGFDGGSHAPVODSRLSDNSAERAEATHLAHSAIRKEAPLAPGNF 505

Qy 1303 TKQIVEKYACTTRISNPNTSQQNFVQTSKRALKQFLPLEETEELKRIITVDSTQWSKN 1362

Db 506 T-----GPCRSNAPRRVKQLKQLIELPLEEIKPEGGVNLNATSEWS-- 548

Qy 1363 MKHLTPSTLTQIDYNEKEKA-TQSPLSCLTRSHSIPQANSFLPIAKVSSFPSPRIPIY 1422

Db 549 ----- 548

Qy 1423 LTRVLPQDNSSHLPAASRYKKKSCVOBSSHFGLOAKKNNLSLAITLLEMTGDQREVGSIG 1482

Db 549 -----ESSPILOAKKNNLSLFLTLEMGAGGQKISALG 582

Qy 1483 TSATNSVYKKVENTVLPKFDLPKTSQKVELLPKVHIYQKDLFPFTSTNSGSPGHLDLVEG 1542

Db 583 KSAAGPLASGKLEKAVLSAGLSEASGAEFLPKVRVREDLLLPQKTSNVSCHAGDLQCE 642

Qy 1543 SLQCTGEGALKWNEANPPGKVPFLRVATESAKTPSKLLDPLDWNHNYGTQIPKEWKSQ 1602

Db 643 IFLQKTRGPNVLNKNRPPG-----RTPSKLLGP-----PMPK-EWESL 679

Qy 1603 EKSPKTAFAKKKDTI-LSLNACSNHIAAINEQNKPEIEVTWAKQRTERLCSQPPV 1661

Db 680 EKSPKSTALRTKDIISLPLDRHESNHSIAAKNEQATQREAAWTQGGPGRUGAPKPPV 739

Qy 1662 LKRQRERITRTTQSDQEEIDYDTTISVMKKEDFDYDSDENOSPRSFKKTRHYFIAA 1721

Db 740 LRRHQDISLPTROPEDKMDYDDIFSTETKGEDFDYDSDENQDPRSPQKTRHYFIAA 799

Qy 1722 VERLWDYGMSSSPHVLNRAQSSGVPQKVVFOEFTDGSFTQPLYSGLNEHLGLGPPY 1781

Db 800 VEQLWDYGMSSSPRALNRRAQNGEVPFRKVVFRERADGFTNPSYRGLNKHGLGLGPPY 859

Qy 1782 IRAEVEDNIM 1791

Db 860 IRAEVEDNIM 869

RESULT 8

KUHU

ferroxidase (EC 1.16.3.1; precursor [validated] - human

N:Alternate names: ceruloplasmin

N:Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)

C:Species: Homo sapiens (man)

C:Date: 31-Aug-1980 #sequence_revision 12-May-1995 #text_change 08-Dec-2000

C:Accession: A25443; A24165; A35450; A00524; I59067

R:Koschinsky, M.L.; Funk, W.D.; van Oost, B.A.; Macgillivray, R.T.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986

A:Title: Complete cDNA sequence of human ceruloplasmin.

A:Reference number: A25443; MUID:86259737; PMID:2873574

A:Accession: A25443

A:Molecule type: mRNA

A:Residues: 1-1060, 1065-1069 <KOS>

A:Cross-references: GB:M3369; NID:g180255; PIDN:AA451976.1; PID:g180256

A>Note: this is the short or CP-2 alternatively spliced form

R:Mercer, J.F.B.; Grimes, A.

FEBS Lett. 203, 185-190, 1986

A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal lead.

A:Reference number: A24165; MUID:86275241; PMID:3755405

A:Accession: A24165

A:Molecule type: mRNA

A:Residues: 1-40; 549-599; 784-829; 919-952 <NER>

R:Yang, P.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, W.A.

J. Biol. Chem. 265, 10780-10785, 1990

A:Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by alte.

A:Reference number: A35450; MUID:90285218; PMID:2355023

A:Accession: A35450

A:Molecule type: DNA

A:Residues: 1007-1064 <YAN>

A:Cross-references: GB:J05506

A>Note: this is the long or CP-1 alternatively spliced form

R:Takahashi, N.; Ortel, T.L.; Putnam, F.W.

Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984

A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequence

A:Reference number: A00524; MUID:84119493; PMID:6582496

A:Accession: A00524

A:Molecule type: protein

A:Residues: 20-1060, 1065-1069 <TAK>

A>Note: 79-Gly and 449-Gly were also found

R:Yang, P.; Naylor, S.L.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McGill

Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A>Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A:Reference number: 159067; MUID:86205676; PMID:13456416
A:Accession: 159067
A:Status: translated from GB/EMBL/DBEJ
A:Molecule type: mRNA
A:Residues: 218-1069 <RES>
A:Cross-references: GB:M13536; NID:g180248; PIDN:AA51975.1; PID:g180249
C:Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions per molecule. In Wilson's disease the plasma levels of ferroxidase are diminished or undetectable. The three fragment chains are produced spontaneously during purification and are genetically distinct.
A:Gene: GDB:CP
A:Cross-references: GDB:1119069; OMIM:117700
A:Map position: 3q23-3q25
A:Introns: 1006/3; 1061/2
A>Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the reduction of NADPH to NADP+.
A>Note: iron(III), but not iron(II), is the form bound and transported by transferrin and lactoferrin.
A:Note: other possible functions are amine oxidase activity, copper transport and homeostasis.
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidoreductase; ferroxidase; ferroxidase repeat homology
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidoreductase; ferroxidase; ferroxidase repeat homology
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1069/Product: ferroxidase, short form #status experimental <MATS>
F:20-1060/Product: ferroxidase, short form #status experimental <MATS>
F:20-493/Product: ferroxidase 67k chain #status experimental <K67>
F:23-357/Domain: ferroxidase repeat homology <FOL>
F:373-718/Domain: ferroxidase repeat homology <FOL>
F:501-905/Product: ferroxidase 50K chain #status experimental <K50>
F:733-1059/Domain: ferroxidase repeat homology <FOL>
F:907-1065/Product: ferroxidase 19K chain #status experimental <K19>
F:138,397,762/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:174-200,276-357,534-560,637-718,874-900/Disulfide bonds: #status predicted
F:227,338,926/Binding site: carbohydrate (Asn) (covalent) #status absent
F:295,338,343/Binding site: copper (His, Cys, His) (type 1) #status predicted
F:358/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental
F:656,699,704,709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F:994,1040,1045,1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 10.7%; Score 1326; DB 1; Length 1069;

Best Local Similarity 19.2%; Pred. No. 2.1e-61;

Matches 399; Conservative 204; Mismatches 419; Indels 1054; Gaps 29;

QY 5 LSTCFCLLRFCFATRYLGLAVELGWYQSDIGE---LPVDARPPRPVKKFPFNT 61
DB 6 LGIFLFCSTP-AWAKKHYYIGIETIWDY-ASDHGKLLSVDEHSNIYQLQGPRI 63
QY 62 SYVYKKTLLFVEETHLNFIAKPRPPMGLLGTIIQAEVYDVTVTITLKNASHPVSLHVG 121
DB 64 GRLYKKALVLYQVTDFTTIEKPVMLGFLGPIIKAEITGDKVYVHLKMLASRPYTFHSHG 123
QY 122 VSYWZASGAEYDDTOSREKEDDKVPFGGSIYVQVVKENGPMASDPLCLTYSLSHV 181
DB 124 ITYKHEGALYDNTDTPQRAADKVPGEQTYMLLATEEESQSGEGGNCVTRIYHSHI 183
QY 182 DAVKDLNSGLICALLVCRBGSIAKEQTQL-HKFTLLFAVFDGKSWHSE-----TKNS 234
DB 184 DAPKIASGLIGLICKDSDKEKHIDREFVYFVSDVNFVSWYEDNFKTYCSEP 243
QY 235 LMQRDAASARAPKMTYGVNYSRLPLGLICHKRSVYVHVGMTTPEVHSIFLEHT 294
DB 244 EKVDKDNEDFQESNRYSVNGYTFGSLPLGLSMAEDREVKWFLFGMGNEVDVEAAPFGQA 303
QY 295 FLVNRHQRASLSIPITFLTAQTLMDLQGLLSCHISQHDGMEAYVYKDSCEEPQL 354
DB 304 LTNKRYRIDTINLFPATLFDAYVAQNPGEWMLSCQNLNHLKAGLQAFQVQEC----- 357
QY 355 RMKNKEAEYDDDLTSEMVDVRFDDNSPSFIQKRSVAKKPKTKTWIYIAAEEDWDY 414
DB 358 ---NKSSKQ-----NTRGKVRH-----YVIAAEELIWNV 385
QY 415 APL-----VLAPDDRYSKSYQLNNGFQRIGRKYKVRFMAYTDETF---XTREAIQ 462

DB 386 APSGIDIFTKENTAPGSDS--AVFPEQGTTRIGGSKYKLVREYTDASTNKRGPPE 443
QY 463 HESGILGLLYGVGVSTLLIIFKQASRPNIYPHGI-----TDVRLY---SRRLP 511
DB 444 EHLGILGPVIAEVDGTVIRVTFNKAGYPLSIEPIGVRFNKNNEGTYYSNYPQSRSP 503
QY 512 KGVKHLKDFPILPGEIFKFKWTVTVEDGPTKSDPRCLTRYYSFVNMERDASGLIGPLL 571
DB 504 PSASH-----VATEFTYVTVKVPINADPVCLAKMYSAVDPTKIDFTGLIGEMK 558
QY 572 ICYKESVDQRGNQIMSDKRVILFSDENRSWYLTENQRFLENPAGVQLEDPPOASN 631
DB 559 ICKKGLSHANGROKVDKEFLPTVDFENESLLEENIRMTTAPQVQKEDDFOESN 618
QY 632 IMHSINGYVDFSILQ-LSVCLHEVAYWILSIGAOTDPLSVFSGSYFKHVMVEDITLTF 690
DB 619 KMTSMGFMYGNGQFLTMCKGDSVWYLFSGAGNEADVHGIFYSGNTYLMWGERDITANLF 678
QY 691 PFGSETVFMGMENPGLWILGCHNSDFENRGMTALLKVSSCDKNTGYEDSYEDISAYLL 750
DB 679 PQTSLTKMPDTEGTFNVECLTTHVTGCKQKYTVNQC----- 718
QY 751 SKNNAIBPRFSQNSRHPSTRQXQFNATTIPENDIEKTDWFAHRTPMPKIQNVSSDLL 810
DB 719 ----- 718
QY 811 MLRQSTPHGLSLSLDQAEKYEFTFSDPSGATDSNNSLSEMTHERPQLHHSQDMVFTP 870
DB 719 --RQSE-----DSTFYLGERTYI----- 735
QY 871 ESGQLRLNEKLGTAATELKLDFKVSSTNNLISTIPSDNLAAGTNTSSLGPPMPV 930
DB 736 -----AAVEVE----- 742
QY 931 HYDSQDITLFGKXSPILTSFSGGPLSSENNDSKLLSGLMNSQESGKNSVSTESGR 990
DB 743 ----- 742
QY 991 LFKGRAHPALLTKDNALFKVISLTKNTKTSNNSATNRKTHIDGPSLLIENSPPVQW 1050
DB 743 -----WD- 744
QY 1051 ILESDETFKVTPLIHDRMLMKDNATLRNLNFMGNKTTSSKNMEMVOOKKEGPIPPDAQN 1110
DB 745 ----- 744
QY 1111 PDMSFFMFLPESARWIQTHGKNSLNSGOGSPKQLVSLGPEKSVGQNFLEKKNVV 1170
DB 745 ----- 744
QY 1171 VGKGEFTKDVGLKEWFPSSRNFLTNLDNLNHNTHNQEKKIQEEIEKETLIQENVVL 1230
DB 745 ----- 744
QY 1231 PQIHTVGTGKNFMKNLFLLSRQNVESYDQYAPVLQDFRSLANDSTNRTKKHTAHFSKK 1290
DB 745 -----YSP----- 747
QY 1291 GEENLEGLGNQTKQIVKYEACTTRISPTSQQNFVTOESKRAUKQFPLPLEETELEKRI 1350
DB 748 -----OR----- 749
QY 1351 IVDDTSTOWSKNMKHLTSTLTQIDYNEKKGAITQSPSLDCLTRSHSIPQANRSPLEIA 1410
DB 750 -----ZWEKELHHL----- 758
QY 1411 KVSPPSIRPLYTRVLFPQDNSSLPAASRYKKOSGVQESSHFLOGAKKNLSAILTLE 1470
DB 759 -----QEQNVNAFL----- 768
QY 1471 MTGDQREVGSLGTSATNSVTKYKVENTVLPKEDLPKTSKVELLPKVHIYQKDLPTETS 1530

```
Db 769 ----- 768
QY 1531 NSPSOHLDLVEGSLLOGTGGAIKWNEANRPKVPFLRVATTESSAKTPSKLLDPLANDHY 1590
Db 769 ----- 768
QY 1591 GTQIPKEEMKSOEKSEKTAFAKKKOTILSINACESNHAIAINEGONKPEIEVTWAKQGR 1650
Db 769 ----- 768
QY 1651 TERLCSQNPVLKRHOREITRTTLOSDBEIDYDDTISVEMKEDFDIYDEBNQSPRSF 1710
Db 769 ----- 772
QY 1711 QKTRHYFTIAVERLWDYGMSSPHVLRNRAQSGVSPQKVVQFQFTDGSFTQPLVRGE 1770
Db 773 ----- 800
QY 1771 LNEHLGLLPYTRAEVEDNIMVTRNQAGRPYSFYSGLISYEDQRCQGAEPKRNFKV--- 1827
Db 801 ESEHLGILQPLHADVDGDKVKKIIFKNWATPEYIHA-----HGVTSSSTVTPFL 850
QY 1828 PNETKYFWKVOHMAKTDKDFCKAWAIPSDVDLEKDVHSGLIGLPLVCHTWTLPAGH 1887
Db 851 PGETLYWVKIPERSGAGTSDSACIPWAVYSTVDQVKDLYSGLIGLPLVCRRLKVFNP 910
QY 1888 RQVTVQEFALFTTIFQETKSWYFTENNERNCRAPCNIQMEDPTKENYRFAHNGYIMDT 1947
Db 911 RRKL--EPALLFLVFDENESWYLDNMTKTYSDHPEKVNKDDEBFISNKKHAINGRMFGN 968
QY 1948 LPLGVMAQDQRIWVLLSGNSNENIHSIHFSGHVFTVRKKEEYKALYNLYPGVFTFEM 2007
Db 969 LQGLTMEVGDENVWYLMGMNEIDLTVHVEHGHFSQYKHGKVSSDVDFIPCTYQITLEM 1028
QY 2008 LPSKAGIWRVECLIGHLHAGMGTFLVYSNKKQTP 2043
Db 1029 FPRTPGTLWLFCHVTDHIAHAGMETTVTLQNEGEYP 1054

RESULT 9
A35210
ferrooxidase (BC 1.16.3.1) precursor - rat
N;Alternate names: ceruloplasmin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A35210, A41753, A29564; S21692
R;Fleming, R.B.; Gitlin, J.D.
J. Biol. Chem. 265, 7701-7707, 1990
A;Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene exp
A;Reference number: A35210; MUID:90237081; PMID:2332446
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1059 <PLE>
A;Cross-references: GB:J05424
R;Fleming, R.B.; Gitlin, J.D.
J. Biol. Chem. 267, 479-486, 1992
A;Title: Structural and functional analysis of the 5'-flanking region of the rat cerulop
A;Reference number: A41753; MUID:92112697; PMID:1730611
A;Accession: A41753
A;Molecule type: DNA
A;Residues: 1-48 <FLZ>
A;Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 a
R;Aldred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.
J. Biol. Chem. 262, 2875-2878, 1987
A;Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plex
A;Reference number: A29564; MUID:87137545; PMID:3818625
A;Accession: A29564
A;Molecule type: mRNA
A;Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'F', 224-226, 'E', 228, 'LL', 231, 'D', 233-235, 'RY
A;Experimental source: liver
A;Note: the authors translated the codon GCG for residue 60 as Gly and GTG for residue 1
R;Ryan, T.P.; Grover, T.A.; Aust, S.D.
```

```
Arch. Biochem. Biophys. 293, 1-8, 1992
A;Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human c
A;Reference number: S21692; MUID:92117681; PMID:1531003
A;Accession: S21692
A;Molecule type: protein
A;Residues: 20-29, 'Q', 902-910 <RYA>
C;Superfamily: ferroxidase; ferroxidase repeat homology
C;Keywords: copper; glycoprotein; oxidoreductase; plasma
F;1-19/Domain: signal sequence status predicted <SIG>
F;20-1059/Product: ferroxidase status predicted <MAT>
F;23-356/Domain: ferroxidase repeat homology <FOA>
F;372-712/Domain: ferroxidase repeat homology <FO2>
F;727-1053/Domain: ferroxidase repeat homology <FO3>
```

Query Match 10.3%; Score 1276; DB 1; Length 1059;

Best Local Similarity 19.1%; Pred. No. 8.7e-59;

Matches 395; Conservative 196; Mismatches 424; Indels 1054; Gaps 29;

```
QY 5 LSTCPFLCLLRFCFSATRRYYLGAVELSWDYMQ-SDLGEL-PVDARFPRVRPKSPFFNTS 62
Db 6 LSALLFL-HSSLAWTRKHHYIGITEAVWDYASGSEKELISVDTQSNFYLRNGPDRIG 64
QY 63 VYKKTFLVEFTDHLFNIAKPPPMGLLGPLOAEVYDVITTKMASHVPVLEAVGV 122
Db 65 RYKKALEYSEYTDGFTXTIDKPAWLGFLGPVKAEGVDKVSVHVKNFASRPYTFHAGV 124
QY 123 SYWKASEGAEYDDQTSOREKEDDKVPPGGSHYVVOVLAKENGPMSDPLCLTYLSHYD 182
Db 125 TYTKANEALYPDNTDFQADKLFPGGQYLVLVLA-NEPSPGSGDSNCVTRIYHSHVD 183
QY 183 LVKDLNSGLIGALLVCREGLSLAKBKQTTL-HKFIILFAVFDEGKSWHSETKSLM----- 236
Db 184 APKDIASGLIGLILLCKGSLHKEENIDQBFVLMFVVDENLSWYLEDNIKTFGSEPE 243
QY 237 -QDRDAASARAPKMTVNGYVNRSLPGLGCHKRKSVMVHVGMTTPEVHSLFLEGHNTF 295
Db 244 KVDKNEDEFOENRMYSINGYTFGSLPGLSMCAEDRVKWLFGMGNEVDVHSLFHGQAL 303
QY 296 LVNRHRQASLEISPTFLTAQTLMDLGOFLASCHTSSHQHDGMEAYKVVDSCFEPPQLR 355
Db 304 TSKVHTDINLPATLIDVSMVAQNPQVWMLSCQNLHLKAGLQAFQVRDC----- 356
QY 356 MKQNEBAEDYDDLDTSEMVDVVRFDNDSFSTQIRSVAKKPKTWHYIAAREEDWDYA 415
Db 357 --NKESP-DDDIQKRHV-----RH-----YYIAAEETIWDYA 385
QY 416 P-----LVLPADDRSYKSOYLNNGPQIRGRKKVRYMAYTDTF---KTRAI 461
Db 386 PSCDTFTGENTLSLGSDBRVFPEQ---GATHIGSYKLVNRYETDSDFTWKEKRGPD 441
QY 462 QHESGILGPLNGEVDGDTLLIIFKNOASRFYNIYPHGTIDVR---PLYSRRRLPKGVKHL 517
Db 442 EEHLGILGPVIMAEVGDILRVTFHNKQFPLSIQPMGVRFTEKNEGTYG---PDGRSSK 498
QY 518 KDFPILPGEILPKYKWTVDGDTKSDPPCLIFYYYSFVNMRERDLASGLIGLLICVKES 577
Db 499 QASHVAPKETFTYEWTPKEMGPTYADPCLSKWYSGVDLTKDIFTGLIGPMKICKGS 558
QY 578 VDQRGNQIMSDKENVILFSVFDNRSWYLTENIQRFLENPAGVQLEDEPEFQASNMHSIN 637
Db 559 ILADGRQKQVDKEFYLPATVFDNENSLLLDDNIMTITAPENVVDKEDDEFOENSKHSMN 618
QY 638 GYVFDLSIQ-LSVCLHEVAYWYILISGAQTDFLSVFFSGVTFKHKVYVEDLTLPFPGSET 696
Db 619 GPMYGNLPLGNMCLGESIYWLFSAGNEADVHGIIYFSGNTYLSKGBRRDTANLPFKSLT 678
QY 697 VFMSMENPGLWILGCHNSDFRNGMTALLKVSCKDNKTGDYEDSYEDISAYLLSKNNAI 756
Db 679 LLMTPTDEGSDVECLTDTDYTGMMKQKYTNOC-----KQGFEDVT- 720
QY 757 EPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFWAHRTMPKIQNVSSDILLMLRQS 816
Db 721 ----- 720
```

QY 817 PTPHGLSLDLOEAKYETFSDDPSGALDENNSLSEMTFRPQLHSGDMVFTPESGLOL 876
Db -----
721 -----
QY 877 RLNEKLGTTAATLKLDFKVSSTSNKLI STIPSDNLAAGTNDTSSLGPPSM?VHYDSQL 936
Db -----
721 -----
QY 937 DTTLFGKKSPFTSGGFLSSENNDSKLLSGLMNSQSSWGNVSSSTESGRLFKGR 996
Db -----
721 -----
QY 997 AHGPAALTCKNALPKVSISSLKTKNTSNNATNRKTHIDGPSLLIENSPPSVKQNILEDST 1056
Db -----
727 TYYYTAA ----- 732
QY 1057 BFKVTVPLIHDRMLDKNATRLNHSNKTSSKNMEMVQKKRGGPIPPDAQNPDMSTFF 1116
Db -----
733 -----
QY 1117 KMLFJPEARWIQTHGKNSLNSGQSPKOLVSLGPEKSVGGQFLSEKKNVVGKGEF 1176
Db -----
733 -----
QY 1177 TKDVGLEKEMVPPSSRNFLFTNLNLHFNTHNQEKIQEEIEKKETLQENWVLPQIHTV 1236
Db -----
733 -----
QY 1237 TGTKNFMKNLFLSTRQNVESYGAYAPVLQDFRSLNDSNTRTKHTAHFSKKEENL 1296
Db -----
733 -----
QY 1297 EBLGNTQKIVEKYACTTIRISNTSQQNFVQTSKRALQFLPLETELEBKIIIVDDTS 1356
Db -----
742 -----
QY 1357 TQWSKNMGLTPSLTLQIDYNEKKGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSSFP 1416
Db -----
743 RDWEMELHHL ----- 752
QY 1417 SRRPIYLRVLFQDNSSHLPAASYRKQSGVOESSHFLQAKKNLSJAILFLEMTGQOR 1476
Db -----
753 -----
QY 1477 EYVSLGTSATNSVYKVENTVLPKDLPLKTSKVELLPKVHIYOKDLFTTSTNSGPH 1536
Db -----
763 -----
QY 1537 LDLVEGSLLOQTGAIKWEANRPGKVPFLRVATSSAKTPSKLLDPLAWDNHYGTQIPK 1596
Db -----
763 -----
QY 1597 BEWKSQKSPKTAFAKKKDTILSLNACSNHAIANEQNKPEIEVTVAKQGRTERLCS 1656
Db -----
763 -----
QY 1657 QNPPVLKRQRBITRTLOSDEIDYDDTISVEMKKEDFDIVDEDNQSPSFQKKTRH 1716
Db -----
763 -----
QY 1717 YFIAAVERLDYGMSSPHVLRNRAQSGVPQFKYVFOEFTDGSFTQPLRYGELNEHLG 1776
Db -----
767 FRIGS -----
QY 1777 LGLPYTAEVEDNIMVTFRQAORPVSFYSSLLISYEDQROGAEPKRFVK---PNETKT 1833
Db -----
801 ILGLPIHADVGAKVYVFKNMATRPYSIHA-----HGVTKSSTVAPTLPGEVRT 850
QY 1834 YFKVQCHHMAPTKDEPDCKAWAYFSDVLEKOVHSGLIGLPLVC---HNTNLNPAHGRQV 1890
Db -----
851 YIWOIPERSGAGTEDSPCPWAYISTIVDRVKDLISGLIGPLIVCRKSYVKVFNPK----- 905

QY 1891 TVQEFALFTIPDETBSWYFTENMERNCRAPCNQIOMEDPTFKENYRPHAINGYIMDTLPG 1950
Db -----
506 KNEFSLLFLVFDENSWLLDNNINYDPDPEKVNKNDEEFISNKKHAIKMGKFGNLQ 965
QY -951 LVMAQDQRIWYLLSGNSNENIHSIHPSGHVFTVRKKEEYKMALYNLYPGVFSTVEMLPS 2010
Db -----
966 LTHVGVGVNWWYVAMGNEIDLTHVPHGHSFQKXRGHISDVDFDFPGTYOTLEMPQ 1025
QY 2011 KAGIWRVECLIGELHAGMSTFLVYSNK 2039
Db -----
1026 TPGTWLHCHVTDHIRHAGVTTTVLFPNQ 1054
RESULT 10
A44258
factor VIII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258
R:Levinson, B.; Kenwright, S.; Camel, P.; Fisher, K.; Gitschier, J.
Genomics 14, 585-589, 1992
A:Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; MUID:93052386; PMID:1427887
A:Accession: A44258
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
A:Cross-references: Q5:W90707; NID:G182316; PID:AAA58466.1; PID:G182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
F1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domain: discoidin I amino-terminal homology <DN2>
Query Match 8.9%; Score 1104; DB 2; Length 216;
Best Local Similarity 100.0%; Pred. No. 6; 7e-51;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2144 VFGNVDSGSKHNPPIIARIYIRLPHYSIRSLRMLMGLMGLDNLSCSMPLGESKA 2203
Db 9 VFGNVDSGSKHNPPIIARIYIRLPHYSIRSLRMLMGLMGLDNLSCSMPLGESKA 68
QY 2204 ISDAQITASSYFTNMFATWSPSKARLHLQGRSNARPOVNNPKWLOVDFOKTKVTGVT 2263
Db 69 ISDAQITASSYFTNMFATWSPSKARLHLQGRSNARPOVNNPKWLOVDFOKTKVTGVT 128
QY 2264 TQGVKSJLTSVMYKBEFLISSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPLLTR 2323
Db 129 TQGVKSJLTSVMYKBEFLISSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPLLTR 188
QY 2324 YLRHPOQSVWHQIALRMEVLGCEAODLY 2351
Db 189 YLRHPOQSVWHQIALRMEVLGCEAODLY 216
RESULT 11
JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl-Cd3 ganglioside
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C:Accession: JC4915
R:Gura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Cross-references: DBJ:U84068; NID:G1620006; PID:BAAL2210.1; PID:G1620007
A:Reference number: JC4915; MUID:96374422; PMID:8780713
A:Accession: JC4915
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DBJ:U84068; NID:G1620006; PID:BAAL2210.1; PID:G1620007
C:Experimental source: CST cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sialic acid.
C:Genetics:
A:Gene: ags

A>Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal an
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-401 <AOR>
R:Mathar, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A>Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
II-like sequences.
A:Reference number: A48394; MUID:93250576; PMID:8485470
A:Accession: G48394
A>Status: preliminary
A:Molecule type: protein
A:Residues: 207-220 <MAT>
A:Experimental source: milk
A>Note: sequence extracted from NCBI backbone (NCBIP:131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: glycoprotein
F:1-32/Domain: EGF homology <EG1>
F:40-79/Domain: EGF homology <EG2>
F:82-239/Domain: discoidin I amino-terminal homology <DN1>
F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.1%; Score 635; DB 2; Length 401;
Best Local Similarity 37.9%; Pred. No. 7.1e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGEHLAGMSTLFLVYSNK-----CQPLGMASCHIRDFQITAGQY-- 2062
DB 50 EQCVTDDSHRG--DVFIQVICKPLGVYGHCECTTCTEPLGQGTGAIADSQISASSMELG 107
QY 2063 ----GQWAPKLARLHYSGSINAWST----KEPFSWIKVDLLAPMIHGKTCQARQKSS 2114
DB 108 FNLGQWAPELARLHQTGIWNAWTSNGYDKNP--WIQVNLKMKWMTGVVTCGASRAGA 165
QY 2115 LYISQFIIMYSIDGKKWQTYRGNSTGTLWVFFGNVDSGKIHFNPPILARYIRLHPHT 2174
DB 166 EYLKTFKVAYSTDGRQFQFIQVAGRSQDKIFIGNVNSGLKINLFDTPLEQYVPLVPII 225
QY 2175 YSIRTLRMELMGCDLNSCMPLGMSKAIQITASSYFTN---MPATWSPKARLHL 2231
DB 226 CHRGTCLRFELGCELGCELTGTEPLGKNDTIPNKQITASSYFTWGLSAFSPFPYVARLDN 285
QY 2232 QGRSNARPOVNNKPEWLOVDFOKTMKVTGVTTCQVKSLLTSMYKKEFLISSQDGHQWT 2291
DB 286 QCKFNAWTAQNSASEWLOIDLGSKQKRVITGITQGRDFGHQYVAARVAYGDDGVWT 345
QY 2292 LFFQNG--KVKVFGQNDSTFPVNSLDPPLLTRYLRHPQSWHQAIRMEVLGC 2345
DB 346 EYKDPGASESKIFPGNMNNSHKNIPTETPQARFVRIQPVAVHNRIILRVELGCG 401

RESULT 15
PAS-6/7 protein precursor - bovine
N/Alternative names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8 h
C:Species: Bos primigenius taurus (cattle)
C>Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text_change 04-Nov-2002
C:Accession: S74211; S78114; S24181; S65138; G48394
R:Harregars, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A>Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob
A:Reference number: S74211; MUID:97008954; PMID:8956064
A:Molecule type: mRNA
A:Residues: 1-427 <HVA>
A:Cross-references: EMBL:X91895; NID:g1632778; PIDN:CAA62997.1; PID:g1632779
A:Accession: S78114
A:Molecule type: protein
A:Residues: 19-95;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
R:Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-212, 1992

A>Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
A:Reference number: S23926; MUID:92353107; PMID:1643094
A:Accession: S24181
A:Molecule type: protein
A:Residues: 383-394 <KIM>
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A>Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal an
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-427 <AOK>
R:Mathar, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A>Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
II-like sequences.
A:Reference number: A48394; MUID:93250576; PMID:8485470
A:Accession: G48394
A>Status: preliminary
A:Molecule type: protein
A:Residues: 233-246 <MAT>
A:Experimental source: milk
A>Note: sequence extracted from NCBI backbone (NCBIP:131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-427/Product: PAS-6/7 protein #status experimental <MAT>
F:24-58/Domain: EGF homology <EG1>
F:66-105/Domain: EGF homology <EG2>
F:108-265/Domain: discoidin I amino-terminal homology <DN1>
F:269-427/Domain: discoidin I amino-terminal homology <DN2>
F:24-35;29-47;49-58;66-77;71-94;96-105/Disulfide bonds: #status predicted
F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F:59;227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:109-265;252-256;270-427/Disulfide bonds: #status experimental

Query Match 5.1%; Score 635; DB 2; Length 427;
Best Local Similarity 37.9%; Pred. No. 7.9e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGEHLAGMSTLFLVYSNK-----CQPLGMASCHIRDFQITAGQY-- 2062
DB 76 EQCVTDDSHRG--DVFIQVICKPLGVYGHCECTTCTEPLGQGTGAIADSQISASSMELG 133
QY 2063 ----GQWAPKLARLHYSGSINAWST----KEPFSWIKVDLLAPMIHGKTCQARQKSS 2114
DB 134 FNLGQWAPELARLHQTGIWNAWTSNGYDKNP--WIQVNLKMKWMTGVVTCGASRAGA 191
QY 2115 LYISQFIIMYSIDGKKWQTYRGNSTGTLWVFFGNVDSGKIHFNPPILARYIRLHPHT 2174
DB 192 EYLKTFKVAYSTDGRQFQFIQVAGRSQDKIFIGNVNSGLKINLFDTPLEQYVPLVPII 251
QY 2175 YSIRTLRMELMGCDLNSCMPLGMSKAIQITASSYFTN---MPATWSPKARLHL 2231
DB 252 CHRGTCLRFELGCELGCELTGTEPLGKNDTIPNKQITASSYFTWGLSAFSPFPYVARLDN 311
QY 2232 QGRSNARPOVNNKPEWLOVDFOKTMKVTGVTTCQVKSLLTSMYKKEFLISSQDGHQWT 2291
DB 312 QCKFNAWTAQNSASEWLOIDLGSKQKRVITGITQGRDFGHQYVAARVAYGDDGVWT 371
QY 2292 LFFQNG--KVKVFGQNDSTFPVNSLDPPLLTRYLRHPQSWHQAIRMEVLGC 2345
DB 372 EYKDPGASESKIFPGNMNNSHKNIPTETPQARFVRIQPVAVHNRIILRVELGCG 427

Search completed: April 13, 2004, 14:09:38
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 18 seconds
(without alignments)
6800.941 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MQIELSTCFCLLRFCFSA.....VVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	12410	100.0	2351	FA8_HUMAN	P00451 homo sapien
2	8827	71.1	2319	FA8_MOUSE	Q06194 mus musculus
3	8068	65.0	2133	FA8_PIG	P12263 sus scrofa
4	2822.5	22.7	2224	FA5_HUMAN	P12259 homo sapien
5	2780.5	22.4	2258	FA5_PIG	P12259 homo sapien
6	2747	22.1	2211	FA5_BOVIN	Q981p1 sus scrofa
7	1322	10.6	1065	CERU_HUMAN	Q28107 bos taurus
8	1273	10.3	1059	CERU_RAT	P08450 homo sapien
9	1242	10.0	1062	CERU_MOUSE	P13635 rattus norv
10	663	5.3	427	MFGM_RAT	P70490 rattus norv
11	657	5.3	463	MFGM_MOUSE	P21956 mus musculus
12	650	5.2	409	MFGM_PIG	P79385 sus scrofa
13	635	5.1	427	MFGM_BOVIN	Q95114 bos taurus
14	588	4.7	387	MFGM_HUMAN	Q08431 homo sapien
15	469.5	3.8	931	NRP2_HUMAN	O60462 homo sapien
16	464.5	3.7	925	NRP2_RAT	O35276 rattus norv
17	462.5	3.7	931	NRP2_MOUSE	O35375 mus musculus
18	458.5	3.7	914	NRP1_CHICK	P79795 gallus gall
19	451.5	3.6	922	NRP1_RAT	Q98wj9 rattus norv
20	446.5	3.6	923	NRP1_MOUSE	P79333 mus musculus
21	443	3.6	928	NRP1_XENLA	P28824 xenopus lae
22	429.5	3.5	923	NRP1_HUMAN	O14786 homo sapien
23	306.5	2.5	3133	HMCT_BOMMO	P98092 bombyx mori
24	266	2.1	764	CPX2_MOUSE	Q94215 mus musculus
25	261	2.1	756	CPX2_HUMAN	Q8436 homo sapien
26	260.5	2.1	280	XLR1_FUGRU	Q98w59 fugu rubrip
27	247	2.0	224	XLR1_MOUSE	Q92114 mus musculus
28	243	2.0	224	XLR1_HUMAN	O15537 homo sapien
29	242.5	2.0	3418	BRC2_HUMAN	P51587 homo sapien
30	235.5	1.9	1358	SIR4_YEAST	P11978 saccharomyc
31	224.5	1.8	3329	BRC2_MOUSE	P97929 mus musculus
32	223	1.8	722	CPX1_MOUSE	Q92100 mus musculus
33	216.5	1.8	1928	MYSL_YEAST	P08964 saccharomyc

34 217 1.7 734 1 CPXM_HUMAN
35 217 1.7 1331 1 CTA2_HUMAN
36 217 1.7 1381 1 YBE7_YEAST
37 216.5 1.7 1271 1 Y338_MYCG
38 213 1.7 1420 1 APX_XENLA
39 212 1.7 1957 1 SPOF_SCHPO
40 212 1.7 3924 1 ANK2_HUMAN
41 211 1.7 1284 1 NEX4_DROME
42 209.5 1.7 1177 1 Y307_MYCG
43 209 1.7 2867 1 RBP2_PLAYB
44 208 1.7 1310 1 CTA4_MOUSE
45 206 1.7 1308 1 CTA4_HUMAN

Q965m3 homo sapien
Q98hc6 homo sapien
P34216 saccharomyc
P47580 mycoplasma
Q01613 xenopus lae
Q10411 schizosach
Q01484 homo sapien
Q94887 drosophila
P47549 mycoplasma
Q00799 plasmodium
Q99p47 mus musculus
Q9c0a0 homo sapien

ALIGNMENTS

RESULT 1
FA8_HUMAN
ID FA8_HUMAN STANDARD; PRT; 2351 AA.
AC P00451.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component)
DE (Antihemophilic factor) (AHF).
GN F8 OR F8C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86081164; PubMed=3935400;
RA Truett M.A., Blacker R., Burke R.D., Caput D., Chu C., Dina D.,
RA Hartog K., Kuo C.H., Maslitz F.R., Merryweather J.P., Najarian R.,
RA Pacht C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
RA Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,
RA Nordfang O., Ezban M.;
RT "Characterization of the polypeptide composition of human factor
VIII:C and the nucleotide sequence and expression of the human kidney
cDNA."
RL DNA 4:333-349 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061548; PubMed=6438526;
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,
RA Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wilson K.L.,
RA Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
RT "Expression of active human factor VIII from recombinant DNA clones."
RL Nature 312:330-337 (1984).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061550; PubMed=6438528;
RA Tootle J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
RA Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr B.C.,
RA Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Pass D.N.,
RA Hewick R.M.;
RT "Molecular cloning of a cDNA encoding human antihemophilic factor."
RL Nature 312:342-347 (1984).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93265012; PubMed=1303178;
RA Gitschier J., Wood W.I.;
RT "Sequence of the exon-containing regions of the human factor VIII
gene."
RL Hum. Mol. Genet. 1:199-200 (1992).
RN [5]
RP SEQUENCE OF 2064-2070 FROM N.A.
RA de Water N.S., Williams R., Browett P.J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SULFATION OF TYR-1699.

- RX MEDLINE=91093266; PubMed=1898735;
RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
RA Mertens K., van Mourik J.A.;
RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RT essential for the interaction of factor VIII with von Willebrand
RT factor.";
RL J. Biol. Chem. 266:740-746(1991).
RN [7]
RN SULFATION.
R2 MEDLINE=92207952; PubMed=1554716;
R2 Pittman D.D., Wang J.H., Kaufman R.J.;
RA "Identification and functional importance of tyrosine sulfate
RT residues within recombinant factor VIII.";
RL Biochemistry 31:3315-3325(1992).
RN [8]
RN STRUCTURE BY NMR OF 2322-2343.
R2 MEDLINE=95200924; PubMed=7893714;
RA Gilbert G.B., Baleja J.D.;
RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
RT amphipathic structure as determined by NMR spectroscopy.";
RL Biochemistry 34:3022-3031(1995).
RN [9]
RN REVIEW ON MOLECULAR BASIS OF HEMA.
R2 MEDLINE=91221499; PubMed=1902642;
RA Gitschier J.;
RT "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
RN [10]
RN REVIEW ON MOLECULAR BASIS OF HEMA.
R2 MEDLINE=89088506; PubMed=2491949;
RA White G.C. II, Shoemaker C.B.;
RT "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
RN [11]
RN REVIEW ON MOLECULAR BASIS OF HEMA.
R2 MEDLINE=95245332; PubMed=7728145;
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RL Hum. Mutat. 5:1-22(1995).
RN [12]
RN VARIANT HEMA GLN-2326.
R2 MEDLINE=86235434; PubMed=3012775;
RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
RT "Identification of a missense mutation in the factor VIII gene of a
RT mild hemophilic.";
RL Science 232:1415-1416(1986).
RN [13]
RN VARIANT HEMA PRO-2135.
R2 MEDLINE=88096339; PubMed=3122181;
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
RT "A novel missense mutation in the factor VIII gene identified by
RT analysis of amplified hemophilia DNA sequences.";
RL Nucleic Acids Res. 15:9797-9805(1987).
RN [14]
RN VARIANT HEMA GLN-2228.
R2 MEDLINE=88191889; PubMed=2893855;
RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
RA Kazazian H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the
RT relative mutation rate at CG dinucleotides.";
RL Am. J. Hum. Genet. 42:718-725(1988).
RN [15]
RN VARIANT HEMA GLY-291.
R2 MEDLINE=88220354; PubMed=2835904;
RA Youssoufian H., Wong C., Aronis S., Platakous H., Kazazian H.H. Jr.,
RA Antonarakis S.E.;
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
RT in exon 7 of the factor VIII gene.";
RL Am. J. Hum. Genet. 42:867-871(1988).
RN [16]
RN VARIANT HEMA CYS-1708.
R2 MEDLINE=89274393; PubMed=2499363;
RA O'Brien D.P., Tuddenham E.G.;
RT "Purification and characterization of factor VIII 1,689-Cys: a
RT nonfunctional cofactor occurring in a patient with severe hemophilia
RL Blood 73:2117-2122(1989).
RN [17]
RN VARIANT HEMA CYS-391.
R2 MEDLINE=90001543; PubMed=2506948;
RA Shima M., Ware J., Yoshioke A., Fukui H., Fulcher C.A.;
RT "An arginine to cysteine amino acid substitution at a critical
RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL Blood 74:1612-1617(1989).
RN [18]
RN VARIANT HEMA LEU-189.
R2 MEDLINE=90057680; PubMed=2510835;
RA Chan V., Chan T.K., Tong T.M., Todd D.;
RT "A novel missense mutation in exon 4 of the factor VIII:C gene
RT resulting in moderately severe hemophilia A.";
RL Blood 74:2688-2691(1989).
RN [19]
RN VARIANT HEMA LEU-2326.
R2 MEDLINE=89197216; PubMed=2495245;
RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
RL Hum. Genet. 81:335-338(1989).
RN [20]
RN VARIANT HEMA HIS-391.
R2 MEDLINE=89264602; PubMed=2498882;
RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Fujimaki M., Hoyer L.W.;
RT "Direct characterization of factor VIII in plasma: detection of a
RT mutation altering a thrombin cleavage site
RT (arginine-372-->histidine).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
RN [21]
RN VARIANT HEMA CYS-1708.
R2 MEDLINE=90105723; PubMed=2104766;
RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
RT Cys) in the factor VIII gene of two unrelated patients with
RT cross-reacting material-positive hemophilia A.";
RL Blood 75:384-389(1990).
RN [22]
RN VARIANTS HEMA GLN-2228 AND LEU-2326.
R2 MEDLINE=90123183; PubMed=2105106;
RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
RT "Recurrent mutations and three novel rearrangements in the factor
RT VIII gene of hemophilia A patients of Italian descent.";
RL Blood 75:662-670(1990).
RN [23]
RN VARIANT HEMA CYS-391.
R2 MEDLINE=90329422; PubMed=1973901;
RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
RT "CRM+ hemophilia A due to a missense mutation (372-->Cys) at the
RT internal heavy chain thrombin cleavage site.";
RL Br. J. Haematol. 75:73-77(1990).
RN [24]
RN VARIANTS HEMA PHE-1699 AND CYS-1708.
R2 MEDLINE=90152691; PubMed=2105906;
RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
RA Kazazian H.H., Antonarakis S.E.;
RT "Characterization of mutations in the factor VIII gene by direct
RT sequencing of amplified genomic DNA.";
RL Genomics 6:65-71(1990).
RN [25]
RN VARIANTS HEMA CYS-1728 AND ASP-1941.
R2 MEDLINE=90169988; PubMed=2106480;
RA Travstman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
RA Kazazian H.H.;
RT "Use of denaturing gradient gel electrophoresis to detect point
RT mutations in the factor VIII gene.";

Query Match 100.0%; Score 12410; DB 1; Length 2351;
Best Local Similarity 100.0%; Pred. No 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MQIELSTCFELCLLRFCSATRRYYLGAVE-SWDYMSQDLGELVDAFFPRVPKSPFN 60
DB 1 MQIELSTCFELCLLRFCSATRRYYLGAVELSWDYMSQDLGELVDAFFPRVPKSPFN 60

QY 61 TSVVYKKTLLVEFTDHLFNTAKRPPPMGLLGPTIOAEVVDTVVITLKNWASHVPSLHAV 120
DB 61 TSVVYKKTLLVEFTDHLFNTAKRPPPMGLLGPTIOAEVVDTVVITLKNWASHVPSLHAV 120

QY 121 GVSWKASGAEDYDDQTSQREKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSWKASGAEDYDDQTSQREKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD 240

QY 241 AASARAMPKMTNGYNYNRSLPGLIGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVRNH 300
DB 241 AASARAMPKMTNGYNYNRSLPGLIGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPIITELTAQTLMDLGOFLLSCHLSSHQHDMBAVYKVDSCPBPQLRMKNNE 360
DB 301 ROASLEISPIITELTAQTLMDLGOFLLSCHLSSHQHDMBAVYKVDSCPBPQLRMKNNE 360

QY 361 EADYDDDLTDEMDVVRFDNDNSPSFIQIRSAVAKHPKTWVHYIAAEEDWDYAPLVLA 420
DB 361 EADYDDDLTDEMDVVRFDNDNSPSFIQIRSAVAKHPKTWVHYIAAEEDWDYAPLVLA 420

QY 421 PDRSYSQVLYNNGPQIRGKYYKVRMAVYDTEFKTREAIQHSGLIGPLLYGEGVDTL 480
DB 421 PDRSYSQVLYNNGPQIRGKYYKVRMAVYDTEFKTREAIQHSGLIGPLLYGEGVDTL 480

QY 481 LIIFKQASRPYNIYPHGIIDVRLYRRLLPKGVKHLKDPILPGEIPKYKWTVTVEDGP 540
DB 481 LIIFKQASRPYNIYPHGIIDVRLYRRLLPKGVKHLKDPILPGEIPKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYYSFVNERDLASGLIGPLLYCYKESVDQRNGQIMSDKRNVLFSVFE 600
DB 541 TKSDPRCLTRYYSFVNERDLASGLIGPLLYCYKESVDQRNGQIMSDKRNVLFSVFE 600

QY 601 NRSWYLTENQRFPLNPAGVQLEDPFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILS 660
DB 601 NRSWYLTENQRFPLNPAGVQLEDPFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILS 660

QY 661 IGAQTDLSVFFSGYTFKHKMWYEDTLTLFPFSGETVFMSENENPGLWILGCHNSDPNRG 720
DB 661 IGAQTDLSVFFSGYTFKHKMWYEDTLTLFPFSGETVFMSENENPGLWILGCHNSDPNRG 720

QY 721 MTALLKVSQDKNTGDIYEDSYEDIAYLLSKNAIAEPSPQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVSQDKNTGDIYEDSYEDIAYLLSKNAIAEPSPQNSRHPSTROKQFNATTI 780

QY 781 PENDIEKTDQFAHRTQMPKIQNVSSDLMLLQSPSTPHGLSLSDIQAKEYTFSDDPS 840
DB 781 PENDIEKTDQFAHRTQMPKIQNVSSDLMLLQSPSTPHGLSLSDIQAKEYTFSDDPS 840

QY 841 PGAIDSNLSSEMTFRPOLHSHGDMVFTPSGQLRLNEKLGTTAATELKKLDFKVSST 900
DB 841 PGAIDSNLSSEMTFRPOLHSHGDMVFTPSGQLRLNEKLGTTAATELKKLDFKVSST 900

QY 901 SNNLISITIPSDNLAAQTDNTSSLGPMPVHYDSQDLDTTLFGKKSPLTBSGGPLSLSEE 960
DB 901 SNNLISITIPSDNLAAQTDNTSSLGPMPVHYDSQDLDTTLFGKKSPLTBSGGPLSLSEE 960

QY 961 NNDSKLESGLMNSQESSWGQVSSTSBGRFLFKGRAHGALLTKDNLAKVSIISLLKTN 1020
DB 961 NNDSKLESGLMNSQESSWGQVSSTSBGRFLFKGRAHGALLTKDNLAKVSIISLLKTN 1020
```

```
QY 1021 KTSNSATNEKTHIDGPSLLIENSPSVQNTLSDTEFKKVTPLIHDRMLMDKNATLRL 1080
DB 1021 KTSNSATNEKTHIDGPSLLIENSPSVQNTLSDTEFKKVTPLIHDRMLMDKNATLRL 1080

QY 1081 NHXSNKTTSSKNMEMVOQKKEGPIPPDAQNPDASFFKXMLFLPESARMIQRTGKNSLNSG 1140
DB 1081 NHXSNKTTSSKNMEMVOQKKEGPIPPDAQNPDASFFKXMLFLPESARMIQRTGKNSLNSG 1140

QY 1141 QGSPKQVLGLGPEKSEVEGQNFLESEKKNVVGKGEFTKDVGLKEMVPPSSRNLFLLTNLDN 1200
DB 1141 QGSPKQVLGLGPEKSEVEGQNFLESEKKNVVGKGEFTKDVGLKEMVPPSSRNLFLLTNLDN 1200

QY 1201 LHENNTNQKCKQEBTEKKEKTELIOENVVLPOHTVVTGTKNMKQLFLLSQNVVSGSYD 1260
DB 1201 LHENNTNQKCKQEBTEKKEKTELIOENVVLPOHTVVTGTKNMKQLFLLSQNVVSGSYD 1260

QY 1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKGEBENLEGLGNQTKQIVEKYACTRISPT 1320
DB 1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKGEBENLEGLGNQTKQIVEKYACTRISPT 1320

QY 1321 SQONFVQSRKRALKQFRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTILTQIDYNEKE 1380
DB 1321 SQONFVQSRKRALKQFRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTILTQIDYNEKE 1380

QY 1381 KGAITQSPLDCLTRSHSIPOANRSLPIAKVSPSPSIRPIYLTRVLFDQNSSHLPAASY 1440
DB 1381 KGAITQSPLDCLTRSHSIPOANRSLPIAKVSPSPSIRPIYLTRVLFDQNSSHLPAASY 1440

QY 1441 RKDQSGVOESHSFLQGAKNKLSIALILTEMTQDQREVGSIGTSATNSVTVYKVENTVLP 1500
DB 1441 RKDQSGVOESHSFLQGAKNKLSIALILTEMTQDQREVGSIGTSATNSVTVYKVENTVLP 1500

QY 1501 KPDLPKTSGKVLLPKVHIYQKDLPTTETNSGSPGHLDLVEGSLLOGTREGAIKWNENRP 1560
DB 1501 KPDLPKTSGKVLLPKVHIYQKDLPTTETNSGSPGHLDLVEGSLLOGTREGAIKWNENRP 1560

QY 1561 GKVPFLRVATESAKTPSKLLDPLANDNHVGTQIPKEEWKSQEKSPKTAFFKXKDTILSL 1620
DB 1561 GKVPFLRVATESAKTPSKLLDPLANDNHVGTQIPKEEWKSQEKSPKTAFFKXKDTILSL 1620

QY 1621 NACESNHAIAINEGONKPEIEVTWAKQGTBELCSQNPVPLKXHQREIERTTLQSDQEE 1680
DB 1621 NACESNHAIAINEGONKPEIEVTWAKQGTBELCSQNPVPLKXHQREIERTTLQSDQEE 1680

QY 1681 IDYDDTLISVEMKKEDFDIYDEENQSPRSQKXTRHYFIAAVERLMDYGMSSSPHVLNR 1740
DB 1681 IDYDDTLISVEMKKEDFDIYDEENQSPRSQKXTRHYFIAAVERLMDYGMSSSPHVLNR 1740

QY 1741 AQSGSVPOQKVVVQBEFTDGSFTQPLYRGLNHLGLLGPYIRAEVEDNTMVTFRNQASR 1800
DB 1741 AQSGSVPOQKVVVQBEFTDGSFTQPLYRGLNHLGLLGPYIRAEVEDNTMVTFRNQASR 1800

QY 1801 PYSFYSSLISYEBDQBGAPPRKVPKNETKTYFWKVOHHMAPTQKDEPCKAWAYPSDV 1860
DB 1801 PYSFYSSLISYEBDQBGAPPRKVPKNETKTYFWKVOHHMAPTQKDEPCKAWAYPSDV 1860

QY 1861 DLEKDVHSGHIGLPLLCHNTNTINPAHQRTVQBEFALPFTIPDETQSWYETENNERCRA 1920
DB 1861 DLEKDVHSGHIGLPLLCHNTNTINPAHQRTVQBEFALPFTIPDETQSWYETENNERCRA 1920

QY 1921 PCNIQWEDPTFKENYFPAHNGVIMDTLPGLVMAQOQRIWYLLSMGNSNENIHSIFSGH 1980
DB 1921 PCNIQWEDPTFKENYFPAHNGVIMDTLPGLVMAQOQRIWYLLSMGNSNENIHSIFSGH 1980

QY 1981 VFTVRKKEEYKMALYNLYPCGVFTVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMALYNLYPCGVFTVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040

QY 2041 QTPILGNASHGHIRFOITASQYQGWAPKLARLHYSGSINAWSTKEPPSWTKVLLAPMII 2100
DB 2041 QTPILGNASHGHIRFOITASQYQGWAPKLARLHYSGSINAWSTKEPPSWTKVLLAPMII 2100

QY 2101 HGIKTQGARQKFSLSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGSKHNI FN 2160
```

```

Db 2101 HGKIQGARQFSSLYISQFIIMSLDGKKQTVRENGSTGTLWFFGVDSSGKHNIFN 2160
Qy 2161 PPIIARYIRLPHTHYSIRSTRMELMGLDNLSCNPLGMSKASIDAGITASSYPTNMPA 2220
Db 2162 PPIIARYIRLPHTHYSIRSTRMELMGLDNLSCNPLGMSKASIDAGITASSYPTNMPA 2220
Qy 2221 TWSPSKARLHLQGRNARVPQNNPKEMQLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLQGRNARVPQNNPKEMQLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTPTVNSLDPELTRYRIHPQSWHQAIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTPTVNSLDPELTRYRIHPQSWHQAIALRM 2340
Qy 2341 EVLGCERQDLY 2351
Db 2341 EVLGCERQDLY 2351

RESULT 2
PA8_MOUSE
ID PA8_MOUSE STANDARD; PRT; 2319 AA.
AC Q06194;
DT 01-JUN-1994 (Rel. 29, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
D2 Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR C8 OR 28C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1] _TAXID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93300511; PubMed=8314577;
RA Elder B., Lakich D., Gitschier J.;
RT "Sequence of the murine factor VIII cDNA.";
RL Genomics 16:374-379(1993).
CC -!- FUNCTION: Factor VIII, along with calcium and phospholipid, acts
CC as a cofactor for factor IXa when it converts factor X to the
CC activated form, factor Xa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Found in most tissues.
CC -!- SIMILARITY: Contains 3 P5/8 type A domains.
CC -!- SIMILARITY: Contains 2 P5/8 type C domains.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L05573; AAA37385.1; -.
CC PIR; A47004; A47004.
CC HSSP; P00451; 1CFG.
CC MGD; MGI:88383; F8.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR008972; Cupredoxin.
CC InterPro; IPR000421; FA58 C.
CC InterPro; IPR008979; GalBind_like.
CC Pfam; PF00394; Cu-oxidase; 3.
CC Pfam; PF00754; P5_F8_type_C; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS50022; FA58C_3; 2.
CC PROSITE; PS00379; MULTICOPPER_OXIDASE1; 3.
CC Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW

```

```

KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19
FT CEAIN 20 2319
FT DOMAIN 20 349
FT DOMAIN 20 349
FT DOMAIN 207 349
FT DOMAIN 399 730
FT DOMAIN 399 573
FT DOMAIN 583 730
FT DOMAIN 760 1640
FT DOMAIN 1683 2008
FT DOMAIN 1683 1845
FT DOMAIN 1855 2008
FT DOMAIN 2008 2156
FT DOMAIN 2161 2313
FT SITE 391 392
FT SITE 759 760
FT SITE 1678 1679
FT SITE 1324 1325
FT SITE 1640 1641
FT MOD_RES 367 367
FT MOD_RES 737 737
FT MOD_RES 738 738
FT MOD_RES 742 742
FT MOD_RES 1669 1669
FT MOD_RES 1687 1687
FT DISULFID 173 199
FT DISULFID 547 573
FT DISULFID 1819 1845
FT DISULFID 2008 2156
FT DISULFID 2161 2313
FT CARBOHYD 61 61
FT CARBOHYD 233 233
FT CARBOHYD 259 259
FT CARBOHYD 423 423
FT CARBOHYD 601 601
FT CARBOHYD 880 880
FT CARBOHYD 958 958
FT CARBOHYD 1015 1015
FT CARBOHYD 1022 1022
FT CARBOHYD 1026 1026
FT CARBOHYD 1044 1044
FT CARBOHYD 1076 1076
FT CARBOHYD 1087 1087
FT CARBOHYD 1136 1136
FT CARBOHYD 1161 1161
FT CARBOHYD 1192 1192
FT CARBOHYD 1255 1255
FT CARBOHYD 1268 1268
FT CARBOHYD 1273 1273
FT CARBOHYD 1274 1274
FT CARBOHYD 1302 1302
FT CARBOHYD 1316 1316
FT CARBOHYD 1340 1340
FT CARBOHYD 1378 1378
FT CARBOHYD 1797 1797
FT CARBOHYD 2105 2105
FT SEQUENCE 2319 AA; 266148 MW; PD054DE051DB2A01 CRC64;

```

```

Query Match 71.1%; Score 8827; DB 1; Length 2319;
Best Local Similarity 72.5%; Pred. No. 0;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;

Qy 1 MQIELSTGFFCLLRFCPSATRRYVILGAVELSWDYMQSD-LGELPVDARFPPRPVKSPFF 59
Db 1 MQIALFACFFLSLNFNFCSSATRRYVILGAVELSWNIQSDLSVLHTDGRFLPRMSTSPFF 60
Qy 60 NTSVYKKTLFVEFTDHLFNIAKPRPPMMGLGPTIQAEVVDVTVITLKNMASHPVSLFA 119
Db 61 NTSIMYKKTVFVEYKQDLFNIAKPRPPMMGLGPTIWTVEVHTVITLKNMASHPVSLHA 120
Qy 120 VGVSVWKASGEGAYDDQTSQREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSILS 179

```

Db 121 VGVSYKASEGEVEDQTSQMEKEDDVFPBESHYVWVLKNGPMAZDPPCLTYSYXS 180
Qy 180 HVDLVKDLNSGLIGALLVCRBGSJAKEXTQTLKHFILLFAVDFDEGSKWSHSETKNSLMODR 239
Db 181 HVDLVKDLNSGLIGALLVCKEGLSKERTOMLYQFVLLFAVDFDEGSKWSESTNDSTQSM 240
Qy 240 DAASARAWPKWHTVNGYNNRSLPGLIGCHRSYKWHVIGMGTTPBHVHSIPIEGHTFLVN 299
Db 241 DSASARDWPKWHTVNGYNNRSLPGLIGCHRSYVWHVIGMGTTPBHSIPIEGHTFFVN 300
Qy 300 HQASLESIPITFLTAQTLMDLGOFLLSCHISSHQDGMAYYKVDSCPEEPQURMK-N 358
Db 301 HQASLESIPITFLTAQTLMDLGOFLLSCHISSHQDGMAYYKVDSCPEESQWQKCN 360
Qy 359 NEBAEDYDDTLDSMVVRDDNSPFIQIRSVAKHPTKWTWYIIAABEDMDYAPLV 418
Db 361 NEEMEDYDDLY-SEMDMFTLDYDSSP-FIQIRSVAKYKPTWYIIAABEDMDYAPSV 418
Qy 419 LAPDDSYKSYLNGPQRIKRYKVPMAVDTETETRAIOHESGILGPLLYGEVD 478
Db 419 PDSNGSYKSYLNGPQRIKRYKVPMAVDTETETRAIOHESGILGPLLYGEVD 478
Qy 479 TLLIIFKQASRPYNIYPHGITDAPLYSRRLPGVHKLDPPIPLPGEIFKYKWTIVVED 538
Db 479 TLLIIFKQASRPYNIYPHGITDAPLYSRRLPGVHKLDPPIPLPGEIFKYKWTIVVED 538
Qy 539 GPTKSDPCLTRYYSFVNMERDLASGLIGPLIICYKSVQORGNQINSDKXNVLPSVF 598
Db 539 GPTKSDPCLTRYYSFVNMERDLASGLIGPLIICYKSVQORGNQINSDKXNVLPSVF 598
Qy 599 DENRSWYITENIORLPLNPAQVQLRDPPOFQASNMHSINGVYFVDSLQSLVCLHEVAYWI 658
Db 599 DENRSWYITENIORLPLNPAQVQLRDPPOFQASNMHSINGVYFVDSLQSLVCLHEVAYWI 658
Qy 659 LSIGAQTDFLSYFFSGYTFKHKWYEDTLTLPFFSGETVFKSMENPGLWILGCHNSDFRN 718
Db 659 LSIGAQTDFLSYFFSGYTFKHKWYEDTLTLPFFSGETVFKSMENPGLWILGCHNSDFRN 718
Qy 719 RGMTHALLKVSCKDNTGYEDSYEDISAYLLSKNNAIBRSPFQNSRHPSTROQFNAT 778
Db 719 RGMTHALLKVSCKDNTGYEDSYEDISAYLLSKNNAIBRSPFQNSRHPSTROQFNAT 778
Qy 779 TIPENDIEKTDWFAHRTMPMKIQNVSSDMLMLLRQS-PTPHGLSLDLOAKYETFS 837
Db 779 TIPENDIEKTDWFAHRTMPMKIQNVSSDMLMLLRQS-PTPHGLSLDLOAKYETFS 837
Qy 838 DPSGAIDSNLSLSEMTFRPQLHSGDMVFTFESGLQRLNEKLGTTAATLKKLDPKV 897
Db 838 DPSGAIDSNLSLSEMTFRPQLHSGDMVFTFESGLQRLNEKLGTTAATLKKLDPKV 897
Qy 839 DHPNAIDSNLSEMTFRPQLHSGDMVFTFESGLQRLNEKLGTTAATLKKLDPKV 898
Db 839 DHPNAIDSNLSEMTFRPQLHSGDMVFTFESGLQRLNEKLGTTAATLKKLDPKV 898
Qy 898 SSTSNNLI-STIPSNLAAGTNTSSLGPPSPVHYDSQDQTLTGKSSPLTESGGPLS 956
Db 898 SSTSNNLI-STIPSNLAAGTNTSSLGPPSPVHYDSQDQTLTGKSSPLTESGGPLS 956
Qy 957 LSEENDSKLIPESGLMNSOESWGNVSTESGRIFKGRACHPALLTKDNALFKVSI-SL 1016
Db 957 LSEENDSKLIPESGLMNSOESWGNVSTESGRIFKGRACHPALLTKDNALFKVSI-SL 1016
Qy 1017 LKTNKTSNNAINRKHIDGPELLNSPVSQWQ-ILESDTEFKKVTPLIHDRLMLKXNA 1075
Db 1017 LKTNKTSNNAINRKHIDGPELLNSPVSQWQ-ILESDTEFKKVTPLIHDRLMLKXNA 1075
Qy 1076 TALRLNHNKNTSSKNMHWQKKEGPIPPAQPDMSFFKMLTLPESARWIORTHGN 1135
Db 1076 TALRLNHNKNTSSKNMHWQKKEGPIPPAQPDMSFFKMLTLPESARWIORTHGN 1135
Qy 1136 SLNSGQSPKQVLSLGPESQVGNFSEKKNVYVKGGEFTKQVGLKEMWPPSSRNIFL 1195
Db 1136 SLNSGQSPKQVLSLGPESQVGNFSEKKNVYVKGGEFTKQVGLKEMWPPSSRNIFL 1195
Qy 1196 TWLNDLHNHNTNNOEKIOEBIEKKTETLQENWVLPQIHVVTGTRKFMKQLFLLSRQNV 1255
Db 1196 TWLNDLHNHNTNNOEKIOEBIEKKTETLQENWVLPQIHVVTGTRKFMKQLFLLSRQNV 1255

Db 1198 TTLSNVHENGHRNQEKNIQEBIE-KEALIBKVVLPQVHEATGSKNFKLIDILLCTRQNI 1256
Qy 1256 EGSVDGAYAVPVLQDFRSLNDSTNTRKHTAHPSK--KGEBENLEGLNQHTQIYKVEYACT 1313
Db 1257 --SYEVEVPVLOQITINNNTVQIHMEHFFKRRKDKETNSGLVNKTEBMYKNY--- 1311
Qy 1314 TRISPNISQOQNFVORSKRALKQFRLPLEETELEKRIIVDDTSTQWSKMKHLPSTILTQ 1373
Db 1312 -----FSQKNITQORSKALQOFL-----STQWLKTNCSQCIKQ 1349
Qy 1374 IDYNEKEGAIQTSPDLSDCLTRSHSIPQANRSPPIAKVGSFPPSIRPIYLTRVLFQDNSS 1433
Db 1350 IDHSEKMKKZP-TKSSLSGS-SVIKSTTQTNSSDHIKVTSAFP---PIDLKRSFPQNKFS 1405
Qy 1434 HLPAAZY-----RKXDSGVOSSEHSHLOGAKKNLSLAULTLEMTGDQREVSGISGTATNSV 1499
Db 1406 HVQASSYIYDPKTSRIQESNNFLKTKINNPSLAILPMNMFIDQGRFTSPGKSNNSV 1455
Qy 1490 TYKKVENTVLPKPDLPKTSQKVELLPKVHIYQKDLFPETSTNGSPGHLDLVEGSLLOQTE 1549
Db 1466 TYKKGEMIIPLKPTLPRESKIELLPQVSIQEBILPTETSHSGSPGHNLAKKEVFLQIKI 1525
Qy 1550 GAIKWNENRPGKVPFLRVATESAKTPSKULLDPLANDNHYGTQIPKEEWSQKSPBKT 1609
Db 1526 GPTKWNKAKRHE--SIKGTESKNTSRKLLNEHAWDYHAAQIPKDMWKSCKSPBII 1583
Qy 1610 AFKKOTILSLNACESNHAINEGQNKPELEVWAKQGSTERLCSQNPVILKSHOREI 1669
Db 1584 SIKOEDTILSRPHGNHSHIGA-NEKQNPQRETTWVKQGTQRTCSQIPVLKRHQBEL 1642
Qy 1670 TRTTLQSDOEBIDYDDTISVEMKXEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDY 1729
Db 1643 --SAFQSOEATDYDDALITIE-TIEDFDIYSEDIKQGRSPQKTRHYFIAAVERLWDY 1699
Qy 1730 MSSSEHVLRNPAQSGSVQFKKVPQBPFTDQSGTQPLYRGEINHEHLGLLGYIIAIVEDN 1789
Db 1700 MSTSL-HVLRNRYQSDNVQPKVWVQBPFTDQSGTQPLYRGEINHEHLGLLGYIIAIVEDN 1758
Qy 1790 IMWTFRQASRPYFYSLSIYEBEDORQAGPRKNVKNPNETKTYFKVQVHMAPTKDEF 1849
Db 1759 IMWTFRQASRPYFYSLSIYEBEDORQAGPRKNVKNPNETKTYFKVQVHMAPTKDEF 1817
Qy 1850 DCKAWAYPSVDVLDKDVHSGSLIGPLLVCHTNTLPAHGRQVTVQBFALFTFIDETKSWY 1909
Db 1818 DCKAWAYPSVDVLDKDVHSGSLIGPLLVCHTNTLPAHGRQVTVQBFALFTFIDETKSWY 1877
Qy 1910 PTENMERNCRAPCNIOXEDPTFKENYRPHANINGIMTDLPGLVMAQDORIRWYLLSMGNN 1969
Db 1878 PTENMERNCRAPCNIOXEDPTFKENYRPHANINGIMTDLPGLVMAQDORIRWYLLSMGNN 1937
Qy 1970 ENIRSIHFSGHVFTVRKKEBKALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGM 2029
Db 1938 ENIRSIHFSGHVFTVRKKEBKALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGM 1997
Qy 2030 STLFLVYGNKQCTPLGMASGHIRDQITASGOYQWAPKARLARLYHSGSINAWSTKEPPSW 2089
Db 1998 STLFLVYGNKQCTPLGMASGHIRDQITASGOYQWAPKARLARLYHSGSINAWSTKEPPSW 2057
Qy 2090 IKVDLLAPMIINGIKTQAROKFSLSYISOFIIMVSLDGKKMOTVGRNSTGLVYFRGNV 2149
Db 2058 IKVDLLAPMIINGIKTQAROKFSLSYISOFIIMVSLDGKKMOTVGRNSTGLVYFRGNV 2117
Qy 2150 DSSGKHNIFNPPIIARYIRLHPHYRSTRLRMLMGLNCSGMLPMSKSAISDAQI 2209
Db 2118 DSSGKHNIFNPPIIARYIRLHPHYRSTRLRMLMGLNCSGMLPMSKSAISDAQI 2177
Qy 2210 TASSYFTNMATWSPSKARLHQGRSNARWPCVNNPKBWLQVDFOKTMKVTGVTQGVKS 2269
Db 2178 TASSYFTNMATWSPSKARLHQGRSNARWPCVNNPKBWLQVDFOKTMKVTGVTQGVKS 2237
Qy 2270 LLTSMYKZEFLLISSQDQHWTLFPQNGKVKVQGNQDSFTFPVNSLDPLTRVLRHP 2329
Db 2238 LLTSMYKZEFLLISSQDQHWTLFPQNGKVKVQGNQDSFTFPVNSLDPLTRVLRHP 2297

Qy	2330	QSWVHQIALREVLGCEAQQDY	2351	FT	DOMAIN	20	199	PLASTOCYANIN-LIKE 1.
Db	2238	QWEHQIALREILGCEAQQQY	2319	FT	DOMAIN	207	357	PLASTOCYANIN-LIKE 2.
				FT	DOMAIN	399	730	F5/8 TYPE A 2.
				FT	DOMAIN	399	573	PLASTOCYANIN-LIKE 3.
				FT	DOMAIN	583	730	PLASTOCYANIN-LIKE 4.
				FT	DOMAIN	760	1599	B.
				FT	DOMAIN	1495	1822	F5/8 TYPE A 3.
				FT	DOMAIN	1495	1659	PLASTOCYANIN-LIKE 5.
				FT	DOMAIN	1669	1822	PLASTOCYANIN-LIKE 6.
				FT	DOMAIN	1822	1970	F5/8 TYPE C 1.
				FT	DOMAIN	1975	2127	F5/8 TYPE C 2.
				FT	SITE	391	392	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
				FT	SITE	759	760	CLEAVAGE (ACTIVATION) (BY SIMILARITY).
				FT	SITE	1449	1450	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
				FT	SITE	1490	1491	SULFATION (BY SIMILARITY).
				FT	SITE	737	737	SULFATION (BY SIMILARITY).
				FT	MOD_RES	738	738	SULFATION (BY SIMILARITY).
				FT	MOD_RES	742	742	PROBABLE.
				FT	DISULFID	173	199	PROBABLE.
				FT	DISULFID	547	573	PROBABLE.
				FT	DISULFID	1633	1659	PROBABLE.
				FT	DISULFID	1822	1970	BY SIMILARITY.
				FT	DISULFID	1975	2127	BY SIMILARITY.
				FT	CARBOHYD	233	233	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	601	601	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	929	929	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	985	985	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1025	1025	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1111	1111	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1181	1181	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1208	1208	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1245	1245	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1265	1265	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1335	1335	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1408	1408	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1611	1611	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CARBOHYD	1919	1919	N-LINKED (GLCNAC. . .) (POTENTIAL).
				FT	CONFLICT	713	713	N -> M (IN REF. 2).
				FT	CONFLICT	734	734	I -> T (IN REF. 2).
				FT	CONFLICT	792	792	G -> Q (IN REF. 2).
				FT	CONFLICT	1133	1133	E -> F (IN REF. 2).
				FT	CONFLICT	1191	1191	I -> L (IN REF. 2).
				FT	CONFLICT	1209	1209	R -> F (IN REF. 2).
				FT	CONFLICT	1437	1437	C -> G (IN REF. 2).
				FT	CONFLICT	1456	1456	F -> R (IN REF. 2).
				FT	CONFLICT	1539	1539	F -> R (IN REF. 2).
				FT	CONFLICT	1546	1546	Q -> N (IN REF. 2).
				SQ	SEQUENCE	2133	2133	AA; 239304 MW; 152BBA8997F570DA CRC64;
								Query Match 65.0%; Score 8068; DB 1; Length 2133;
								Best Local Similarity 67.0%; Pred. No. 0;
								Matches 1577; Conservative 208; Mismatches 344; Indels 226; Gaps 17;
Qy	1	MOELSTCFCLLPFCPSATRRYVLGAVELSDWYMQSD-IGELPDVDAFPPRPVKSPPF	59	Qy	1	MOELSTCFCLLPFCPSATRRYVLGAVELSDWYMQSD-IGELPDVDAFPPRPVKSPPF	59	
Db	1	MOELSTCFCLLPFGFSATRRYVLGAVELSDYRQSELLRELVDFRFPATAPCALPEL	60	Db	1	MOELSTCFCLLPFGFSATRRYVLGAVELSDYRQSELLRELVDFRFPATAPCALPEL	60	
Qy	60	NTSVVYKTLVEFTDHLFNIAKPPWMLGPTIOAEVVDVTVITLKNMASHPVSLHA	119	Qy	60	NTSVVYKTLVEFTDHLFNIAKPPWMLGPTIOAEVVDVTVITLKNMASHPVSLHA	119	
Db	61	GESVLYKTVFVEFTDQLFSVARPPPMWMLGPTIOAEVVDVTVITLKNMASHPVSLHA	120	Db	61	GESVLYKTVFVEFTDQLFSVARPPPMWMLGPTIOAEVVDVTVITLKNMASHPVSLHA	120	
Qy	120	GVSVYKKAEGAEYDDQTSQREKEDKVPFGSGSHYVQVLKENGPMASDPLCTYSYLS	179	Qy	120	GVSVYKKAEGAEYDDQTSQREKEDKVPFGSGSHYVQVLKENGPMASDPLCTYSYLS	179	
Db	121	GVSVYKKAEGAEYDDQTSQREKEDKVPFGSGSHYVQVLKENGPMASDPLCTYSYLS	180	Db	121	GVSVYKKAEGAEYDDQTSQREKEDKVPFGSGSHYVQVLKENGPMASDPLCTYSYLS	180	
Qy	180	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKFILLFAVDFDEKSHWSETKNSLMQDR	239	Qy	180	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKFILLFAVDFDEKSHWSETKNSLMQDR	239	
Db	181	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKFILLFAVDFDEKSHWSETKNSLMQDR	240	Db	181	HYDLVKDLNSGLIGALLVCRGSLAKGKTQTLHKFILLFAVDFDEKSHWSETKNSLMQDR	240	
Qy	240	DAASARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPVHSTFLEGHTFLVFN	299	Qy	240	DAASARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPVHSTFLEGHTFLVFN	299	
Db	241	DPAPARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPVHSTFLEGHTFLVFN	300	Db	241	DPAPARAFKMTVNGYVNRSLPGLIGCRKSVYWHVIGMCTTPVHSTFLEGHTFLVFN	300	

QY 300 HQASLEISPIITLTAQTLLMDLQFLLSCHISSHOEDCHBAVVKVDSCEBEPQLEMKN 359
DB 301 HQASLEISPLTFLTAQTFLMDLQFLLFCHSSHGGMEAFVRVESCAEBPQURRAD 360
QY 360 REAEDYDDDLTSEMDVDFDDNSPFIQIRSAVAKKHPTWVHYIAABEEEDWDYAPLV 419
DB 361 EE-EDYDDNLYSDMDVRLDGDVDFPFIQIRSAVAKKHPTWVHYIAABEEEDWDYAPV 419
QY 420 APDRSYKQYLANGPQIRGRKYKVRFWAYTDETFTATREAIQHEGILGLPLLYGBVGD 479
DB 420 SPDSRYSKSYLNSGFCQIRGKYKARFVAYTDVTEKRAKAIPIYESGILGPLLYGBVGD 479
QY 480 LLIIFKQASRPNIYPHGTIDVRPLYSRRLPGVGHLDFFLPGLPGRIKPKYKWTWVEOG 539
DB 480 LLIIFKQASRPNIYPHGTIDVSALHGGRLGKWHLDXMPILPGETFIKPKYKWTWVEOG 539
QY 540 PTKSDPRCLTRYVSSFWNMERDILASGLIGPLLICYKESVDQRGNQMSDKRNILFSVFD 599
DB 540 PTKSDPRCLTRYVSSINLEKDLASGLIGPLLICYKESVDQRGNQMSDKRNILFSVFD 599
QY 600 ENRSWYLTENIQRLPNPAGVOLEDPPFQASNMHSHNGVVFDSQLSVCLHEVAWYIL 659
DB 600 ENQSWYLAENIQRLPNPDGLQDPFQASNMHSHNGVVFDSQLSVCLHEVAWYIL 659
QY 660 SIGAOTDFLSVFGYTFKHVMVVEDTLTFPSGETVMSMENPGIWLILGCHNSDFNR 719
DB 660 SVGAOTDFLSVFGYTFKHVMVVEDTLTFPSGETVMSMENPGIWLILGCHNSDLNR 719
QY 720 GMTALLKVSCKNTGDIYEDSYEDI SAYLLSKNAIEPRSFQSNRHPSTQKQFNATT 779
DB 720 GMTALLKVSCKRDIGDYDNTYEDI PCGLLSGKNVIEPRSFQSNRPPPSASQXQFTIT 779
QY 780 IPENDIKTPWFAPHRTPMKIQNVSSDMLMLRQSPFHGLSLDLOEAKYETPDDP 839
DB 780 SPEDDVE-LDPQSGERTQALEELSVPGSGMLLQNPAPHGSSSDLOEARNE--ADDY 836
QY 840 SPGAIDNSNISSEMTHERPQLHSGDMVFTPESGLOLRNLKJCTTAATELKLDRKVS 899
DB 837 LFGARRENTAPSAARURPELHHAERVLTEP-----EK-----ELKLDKMS 882
QY 900 TSNNLIS--TIPSDNLAAGTDNTSLGPPSPVHYDSQDITFLFKKSSPTESGGLSL 957
DB 883 SSDLLKCSFTIPSDLSAETERTSLGPPHPQVNFPSQIGAIVLGKNSHFGAGVPLGS 942
QY 958 SHENDSKLESGLMNSQESGAKNVSTESGRFLPKXRAHGPAALLTKONALPKVUSILL 1017
DB 943 TEEDH-----ESSLGENVSPVESDGIPEKERAHGPAASLTKDDVLFKVNISLV 989
QY 1018 KTNKTSNNSATNRKTHIDGPSLLIENSPPVQWNILESDETFKKVTPLIHDMMLDKNATA 1077
DB 990 KTKARVYLTNRKTHIDDAALLTENRASA-----TFMDKNVTA 1028
QY 1078 LRLNHSNKTTSKKNVEMVQKKEGPIPPDAQNDPMSFFKMLFUPESARWQRTHGKNSL 1137
DB 1029 SGLNHSN-----WIKPLGKNPL 1047
QY 1138 NSGQGPSKQIVSLGPEKSVBQNFLEKKNVVKVKGFTDVGLEKEMVFPSSRNLELTN 1197
DB 1048 SSERGPSPELLTSSGSGSVKSGQSGQGRIRVAVBEEELSKG---KENWLPNSELTFUTN 1104
QY 1198 LDNLHENHNOEKKIQREIEKETLIQENVVLPQIHTVVTGTFNPMKNLFLLSQNVG 1257
DB 1105 SADVQGNTHSGKKSREEMEREXLVQEKVDLPQVYIATGFKNLFNIFHQSTEPSVEG 1164
QY 1258 SYDGAAYAPVQDFRSLNSTNTKHTAHPSKKGBEENLEGNGNOTKQIVEKYACTTRIS 1317
DB 1165 FDCGSHAPVPOQSRGSLNDSARAETHIAHFAAREEAPLEAPCNRT----- 1210
QY 1318 PNTSQCNFTVQSKKALQKFRIPLETELEKLIIVDDTSTQWSKNMKHLPSTLTQIDYN 1377
DB 1211 -QPGFSAVPRVRKQSLKQIRLFELEBEIKPGRGVLNATSTRWS----- 1252

RESULT 4

PAS_HUMAN

ID PAS_HUMAN

AC

STANDARD;

PRT: 2224 AA.

P12259; Q14285;

QY 1378 EXEKGAITQSPLSDCILTRSHSIPQANRSPLIPTAKVSSP2SIRPIYLTRVLFDQNSHLPA 1437
DB 1253 ----- 1252
QY 1438 ASYRKDSQVQSSSHPLQAKAKNNLSAILITLMTGDQREVCSLGTASNTSVYKVENT 1497
DB 1253 -----ESSPIQAGAKNNLSPLFTILEMAGGQGIKISALGSAAGPLASGKLEKA 1301
QY 1498 VLPKPDLPKTSKGVLLPKVHIYOKDLRPTETSNQSPGHLDLVEGSLLOCTEGAIKNFA 1557
DB 1302 VSSAGLSASAKABFLPKVRVREDLLLPQKISNVSCAHGDLQOETFLQKTRGPVNLKV 1361
QY 1558 NRPKGVPFLRVATESSAKTFSKLLDP1AWDNHYGTQIPKEMKSOQSKSEPKTAFKKOTI 1617
DB 1362 NRPK-----RTPSKLLGP-----PMPK-EMESLEKSPKSTALRTKDII 1398
QY 1618 -LSLNACENHAIANBEGONKPEIIVTWAKQGRTERLCSOMPVLKRBORBITRRLQS 1676
DB 1399 SLPLDSESHNSIAAKNEGCAETQREAAWTKQGGPGRLCAPKPPVLRHQDLSLPTFQ 1458
QY 1677 DQEEIDYDDTISVEMKKEDFDIYDEDEENQSPRSFOKKTREHYPIAAVERLWDYGMSSPHV 1736
DB 1459 BEDXMDYDDIFSTETGDEDFDIYGEDENQDPRSFQKTRHYPIAAVEQLWDYGMSSPRA 1518
QY 1737 LRNRAQSGSVQPKKVVFOBFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVTFRN 1796
DB 1519 LRNRAQNGEVPREPKVVREFAFGSFTQPSYRGELNKHGLLGPYIRAEVEDNIMVTFKN 1578
QY 1797 QASRPVSFYSLLSYEEDQCGAEPKRVKNETKTYFWKVOHNVAPTXDEDFCKAWAY 1856
DB 1579 QASRPYSFYSSLLSYDDQCGAEPRHNFVQNETRTYFWKVOHNVAPTXDEDFCKAWAY 1838
QY 1857 FSDVLEKDVHSGLIGPLLCHVTNLTNPAHGRQVTVQEFALFTTIFDETSKSWYFTENMR 1916
DB 1639 FSDVLEKDVHSGLIGPLLICRANTLNAHGRQVTVQEFALFTTIFDETSKSWYFTENMR 1698
QY 1917 NCRAPONIOMEDTFKENYRPHAINGYIMDTLPGVNAQQRIRWYLLSGNSNENIHSIH 1976
DB 1699 NCRAPCHLQMEDPTKENYRPHAINGYIMDTLPGVNAQQRIRWYLLSGNSNENIHSIH 1758
QY 1977 FSGHVFTVRKEEYKMLNLYPGVETVEMLPKAGIWEVECLIGHLLHAGMSTLFLVY 2036
DB 1759 FSGHVFTVRKEEYKMAVYNLPGVETVEMLPKAGIWEVECLIGHLLHAGMSTLFLVY 1818
QY 2037 SNKQOTPLGNASGHIRDFOITASQYQGWAPKLARLHYSINAWSTKEPFSWIKVDLLA 2096
DB 1819 SKQOAPLGNASGHIRDFQITASQYQGWAPKLARLHYSINAWSTKEPFSWIKVDLLA 1878
QY 2097 PMIHGKTQCAROKFSLSYISQIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSSGKIH 2156
DB 1879 PMIHGINTQCAROKFSLSYISQIIMYSLDGRNMQSYRGNSTGTLMVFFGNVDASGKIH 1938
QY 2157 NIENPPIIARVIRLHPHYSIRSTRVELMCDLNSCSMPLGMESKAISDAQITASSYFT 2216
DB 1939 NIFPPIVARVIRLHPHYSIRSTRVELMCDLNSCSMPLGMQNKALSDSQITASSHLS 1998
QY 2217 NMFATWSPSKARLHLOGRSNAWRPQVNNPKWLQVDFQTKMKVGTGVTTCQVKSLTSMYV 2276
DB 1999 NIFATWSPSQARLHLOGRTNAWRPRVSAEELQVDIQTQKVTGITTTCQVKSLTSMYV 2058
QY 2277 KEFLISSQDQHWTLRFQNGKVKVFCQNGQDSFTFVNSLDPPLLTRYLRIHPQSVWHQI 2336
DB 2059 KEFTLVSSQDGRWTLFLQDGHFTKVFQNGQDSSTFVNVNLDPLPLFTRYLRIHPTSWAQHI 2118
QY 2337 ALRVEVLGCEAQLDY 2351
DB 2119 ALRVEVLGCEAQLDY 2133

DT 01-OCT-1989 (Rel. 22, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 DE F5.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A., AND VARIANT MET-1764.
 RX MEDLINE=92232658; PubMed=1567832;
 RA Cripe L.D., Moore K.D., Kane W.H.;
 RA "Structure of the gene for human coagulation factor V";
 RL Biochemistry 31:3777-3785(1992).
 [2]
 RX SEQUENCE FROM N.A., AND VARIANT GLU-925.
 RX MEDLINE=87260886; PubMed=3110773;
 RA Jerry R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.W., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 [3]
 RX SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GLU-925 AND ILE-1285.
 RX MEDLINE=88107560; PubMed=2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 region of human factor V, a blood coagulation factor with four types
 of internal repeats";
 RL Biochemistry 26:6508-6514(1987).
 [4]
 RX SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE=86313663; PubMed=3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 factor homologous to factor VIII and ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 [5]
 RX PARTIAL SEQUENCE FROM N.A.
 RX TISSUE=Endothelial;
 RX MEDLINE=93203619; PubMed=8454869;
 RA Shen N.L.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 lymphocytes";
 RL J. Immunol. 150:2992-3001(1993).
 [6]
 RX SULFATION.
 RX MEDLINE=94264012; PubMed=8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
 RA Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 thrombin cleavage and activation and for full procoagulant activity";
 RL Biochemistry 33:6952-6959(1994).
 [7]
 RX SULFATION.
 RX MEDLINE=90366699; PubMed=2168225;
 RA Hortic G.L.;
 RT "Sulfation of tyrosine residues in coagulation factor V";
 RL Blood 76:945-952(1990).
 [8]
 RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE=20052169; PubMed=10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 coagulation factor V";
 RL Nature 402:434-439(1999).
 [9]
 RX VARIANT MET-1764.
 RX MEDLINE=95179146; PubMed=7874144;
 RA Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;

RT "A polymorphism in the human coagulation factor V gene";
 RL Hum. Mol. Genet. 3:2085-2085(1994).
 [10]
 RX VARIANT APCR GLN-534.
 RX MEDLINE=94217810; PubMed=8164741;
 RA Bertina R.M., Koelman B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 activated protein C";
 RL Nature 369:64-67(1994).
 [11]
 RX VARIANTS ILE-1285 AND ARG-1327.
 RX MEDLINE=96351768; PubMed=8713778;
 RA Lunghi B., Iacoviello L., Gemmati D., Dilasio M.G., Castoldi E.,
 RA Pinotti M., Castaman G., Redaelli R., Mariani G., Marchetti G.,
 RA Bernardi F.;
 RT "Detection of new polymorphic markers in the factor V gene:
 association with factor V levels in plasma";
 RL Thromb. Haemost. 75:45-48(1996).
 [12]
 RX VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RX MEDLINE=98122763; PubMed=9454741;
 RA Chan W.P., Lee C.K., Kwong Y.D., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese";
 RL Blood 91:1135-1139(1998).
 [13]
 RX VARIANT APCR THR-334.
 RX MEDLINE=98122764; PubMed=9454742;
 RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 resistance to activated protein C";
 RL Blood 91:1140-1144(1996).
 [14]
 RX VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;
 RX ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;
 RX ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
 RX MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 RA Lander E.S.;
 RT "Characterization of single-nucleotide polymorphisms in coding regions
 of human genes";
 RL Nat. Genet. 22:231-238(1999).
 [15]
 RX ERRATUM.
 RX Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.Q.,
 RA Lander E.S.;
 RL Nat. Genet. 23:373-373(1999).
 [16]
 RX VARIANT APCR HIS-2102.
 RX MEDLINE=21847288; PubMed=11858490;
 RA Schrijver I., Houissa-Kastally R., Jones C.D., Garcia K.C.,
 RA Zehnder J.L.;
 RT "Novel factor V C2-domain mutation (R2074H) in two families with
 factor V deficiency and bleeding";
 RL Thromb. Haemost. 87:294-299(2002).
 CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
 with factor Xa to activate prothrombin to thrombin.
 CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
 chain, noncovalently bound. The interaction between the two chains
 is calcium-dependent.
 CC -!- DOMAIN: Domain B contains 35 x 9 AA tandem repeats, and 2 x 17 AA
 repeats.
 CC -!- PTM: Thrombin activates factor V proteolytically to the active
 cofactor, factor Va (formation of a heavy chain at the N-
 terminus and a light chain at the C-terminus).
 CC -!- PTM: Sulfation is required for efficient thrombin cleavage and
 activation and for full procoagulant activity.
 CC -!- DISEASE: Defects in F5 are the cause of Owren parahemophilia
 [MIM:227400], an hemorrhagic diathesis.

389 KHYKVVYKQYQDSFTKRLNPNKNGDILGPIRAQVRDTLKIIVFNWASRSYSIVPH 448
498 GIT-----DVRPLYSRLPKGVXHLKDPFPLPGEIKYKWTIVVEDGPTKSDRCLTRY 551
449 GVTFSPVEDDYNSSSTSDNNTMIR-----AVQGETYTYKWNILESDEPTENDAQCLTRP 503
552 YSSFNVERDIASGLIGLPLLCYKESVDQRGNQMSKRNVLFSVFDENRSMYLTENIQ 611
504 YYSNVDIRDIASGLIGLPLLCYKESVDQRGNQMSKRNVLFSVFDENRSMYLTENIQ 563
612 RFLPNACVQLEDEFEQASNMVMSINGVVFBSL-QLSVCLHEVAYWYILSICAQDTFLSV 670
564 KFCNPKVKZDDKFEYSNMSTINGVVPBSIPTLGFCDTQWMEFCSVRTHDNILTI 623
671 FFSGVTPKHMVYEDTTLFPFSGBTVMSENPGWLWILGCHNGDFRNGMTALLKYSSC 730
624 HFTGHSFYKGRHDTTLFPNGBESVTVMONVGTWMLTWNPNRKNKLQKFRDYKC 683
731 DKNTG-DYVEDSYEDISAYLL-----SKNNAIEP-----RSFQNS 765
684 IRDDDEDSYELIYRPSSTTLTRKWHDSSENKEBENDCEYDQDLASVLGIRSFNRS 743
766 RHPSTROQFNATIP-ENDIEKIDPWFAHRTMPKIQNVSSDOLLMLLRQSPHGLSL 824
744 LY--QEDDEFNLALALENNESEFTP-----STD-----RAVDSNSSP 780
825 SLOQAKVETPSDD---PSGAIDSNNS-----LSEWTFRPOLHSHGDMVFP-PSG 873
791 GNISRAPANTTEPRKIILPHFEATKAGSPRHGTGLVNLVLRRTQHSDDPYSEDPIENP 840
874 IQ-----LRLNE-----KLGTAAATE--LKLDPKVYSTSNNLISIP 909
841 LQSVITG-SILPFGTEGRNREHPKHKFAGRDQAAKHRSQMEFFPAHKTGRH----IS 896
910 SDNLAACTN-TSSLGPPMPHYDSLOTLTLGKSSPLTESGGPLSSENN-----962
897 QDN-----SSSSSMGPLE-----DLSDDLLEK-DESTINGKWHLVSEKGSYIYQDA 945
963 DSKILESGLMNSQES---SWGKNVSTBSGRLFKGRAGPALLKDNALFKVSIISLKT 1019
946 DEDMAVKNLPNNFQNASWGENIIFTNK---HGKQGHPIFVTEH-----KLJQE 993
1020 NKTNSNS-----ATNRKTHIDGSPLLIENSPSWQNIL-ESDTEFKVTPLIHDM 1069
994 RQDEGNSILKGRFFIRTRRKKERKPVHIVLSPRSFNLGEANTPFS-----1043
1070 LMDKNATALRNMENKNTSSKNMVMVQKGGP'PPDAQN---PDMGFFKMLP'PESAR 1126
1044 ---DRQNHSLLEHNET-----PPFDLNOTFPMXNLSLASHPD---1082
1127 WIQRTHGKNSLN-SGQGPSPKQIVSLGPKSVEGQNF'LSERKNVYVVGGEFTKDYGLKEM 1185
1083 ---HNQNLNDTHQTSF-----LDLYGT 1103
1186 VFPSS--RNLFTNLNDLHNTHNQEKIQEIEKKEKTELIOENVVLPOIHVVTGKNFM 1243
1104 VTFDEPYQAP'QDLDPETHSTAVPSQSLPEPI-----QXH-----1140
1244 KXNLLISTEQNEGSYDGA'VPIQDPRINDSTNRTKXHTAHS---KKGBEENLEG 1300
1141 -----DYD-----LRKASPTDSEMPFSLKLAGHRTTSPDJ- 1173
1301 NQTKQIVEYACT-----TRISPTNSQNFVQSRKALKQFRLPLETEL-----BK 1348
1174 NQTSLSPELSQTTLSDPGHWTLSPDLSQTTLSPLDLSH-TLSP---DLGHTTSLSPDLSHT 1230
1349 RIIVDTSTQWKNMKHLPST-TQIDYNEKEKAITQSPDLSCLTRSHSIPOANRSLP 1408
1231 TLSPDLSQTTLSPLDLSHTLS-----PDLGHTTSLSPDLSHTTSLSPDLSHT-D 1280
1409 IAKVSSFPSTRIPIVLEVLFDQNSHLPAASYKKDLSG-VQESSHFLQGAKKNNLSIAL 1467
1281 LSHTTLSPLDLSHTTSLSPDLSHTTSLSPDLSHTTSLSPDLSHTTSLSPDLSHTTSL 1340

1468 T--LEMTGDQREYSGSLCTSATNSVTKYKVENTVLP-----KFDLPKTS-----GKVEL 1513
1341 SPDLSSHITLSPDLGHTTSLSPDLSQ-----TLSPDLGHMTLSPDLSETTSLSPDLGHTT 1394
1514 LPKVHIYOKOLFPTETNSGPGHLDLVEGSLLOGTEGAIKWNANRPGKVPFLFVAVTSS 1573
1395 SP-----DLSHT-TLSPDLGHMTLSPD---LGQTTLSLDLFGQTTLSPLDLSHTLSSELS 1444
1574 AKTSPKLL-----DPLAW-----DNHYGTQIPKEEMKQSKSPKTAFAKK 1613
1445 HETLSPDLSQVTLSPDLSSEIPFSPDLMQTTLSLSDNLTTLSPLDLSQSPDPDEPKTSY-- 1502
1614 KDTILSLNACESHAIAINEGQNKPEIETWAKQGSTERLCSQ-----NPPVLKXHQ 1666
1503 -----ISESSQSVLP'EPFGQTSPPDLCQRPSPSHSTLNNTFIPREFNPMVWVGLS 1554
1667 RE-----LTRITLQSDQES-----IDYDDDTISVENKKEDEPIYDEDNQSB----- 1707
1555 RDDGDYVEIIPROQEENSEEDYKIDYVEDDPYQTDVRT-----DINSRNPNDNTAAW 1608
1708 --RSFQKTRHYFTAAVERLWDYQMSSSPHVLRNRAQSGVPO---FKXVVFQFTDGSF 1762
1609 YLRNNGNRNYIARABELSWDYK-----FTQREIDDDVPEHTIYKVVFRKYLDSFT 1662
1763 TQPLYRGELNEHLGLGPIYRAEVEDNIMVTFRNOAGRPYSFYSSLSIYE-----ED 1814
1663 TKLDPGRGEYEHGILGPIIRAEVDVIVQVFRKNLASRPYSLHAHGLSYEKSSSEKTYED 1722
1815 ORQAEERKPNFVKNETKTYFKVCHHNAFKDEFCKANAYFSDVLEKDVHSLGLPL 1874
1723 DSPWFCKEDNAVQNSSTYVYWHATERSGPSPGACRWAYVSAVAFPEKDHSGLGLPL 1782
1875 LVCHTNTLNPAHQSVTVQBPALFPTIDETKSWYFTENMERNCRAPCNIQMEDPTFKEN 1934
1793 LICKGTUHKENMEVDMBEFVLL'FWYFDEKKSWMYKFTESNR-----LTSSEVENS 1836
1935 YRFHAINGYIMDTL'PGLVMAQDQIRWYLLNGNENIHSIHFGSHVTVTKKSEYKAL 1994
1837 HKFETAINGYIN-LPGLFMYEQEWRVRLHLLNLGGSRD'HHVVHFGQTLLENGTQOHLGV 1895
1995 YNLYPGVETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGWSGHIRDF 2054
1896 WPLLPSPKTL'EMKTSKAGWMLLDTVEGENQAGNQTPFLIIDRECKPMGLSGLIADS 1955
2055 QITASQYQGWAPKLARLHYSGSINAWSTKEPFS-----WIKVDLLAPMIHGIKTQG 2107
1956 QIKASEFWGHQPKLARLNNNGSYNAWIT-DKFSGESNSKZWIOVDMQREVVP'GIQTQG 2014
2108 ARQFESSLYISQFTIMYSLDGKQWQTYGNSGTGLIMVFFGNVDSGIXHNTFNPIIARY 2167
2015 AKYLLKSYTTFPNVAYSSDQNRWIRPKGNSTKNWYFNGNSDASTITENOFDPPVARY 2074
2168 IRLHETHYSIRSTRVRLVLMGCDLNSCMPLGWESKAI'DAQITASSYPTNNFAT--WSPSK 2226
2075 IRISTESYNKALRELOGCEVNGCCSTPLGYESNIKNEQITASSFKKSWGDM'WBPFR 2134
2227 ARHLHQGRSNARPOVNNPKWLQVDFOKTMKVTGVTTCQVKSLTSMYVKEFLISSQD 2286
2135 AFLNAQGRVNAWQAKANNNNQWLQIDLLKIKITAITTQGCCKSLSESMKVKRYTIQYSDR 2194
2287 GHQWTLFQNGVY--KVFCGNQDSFTPVNSLDPDLLTYLRIHQSVWVHQAALRMEVLG 2344
2195 GYEMKSYREKSMWDKI'PEGNNNKHGKVNFPNPIISRFIRIIPKWNQSI'ALRELF 2254
2345 CEAQDLY 2351
2255 C---DIY 2258
RESULT 6
FAS_BOVIN
ID_FAS_BOVIN STANDARD; PRT; 2211 AA.

AC Q28107; Q28108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
CX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver.
RC MEDLINE=92147638; PubMed=1737753;
RA Guinot E.R.; Esmen C.T.; Mann K.G.; Macgillivray R.T.;
RT "The complete cDNA sequence of bovine coagulation factor V";
RL J. Biol. Chem. 267:2971-2978(1992).
RN [2]
RP MEDLINE=95034740; PubMed=7947716;
RX Xue J.; Kalaratis M.; Silveira J.R.; Kung C.; Mann K.G.;
RA "Determination of the disulfide bridges in factor Va heavy chain";
RL Biochemistry 33:13109-13116(1994).
CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
CC with factor Xa to activate prothrombin to thrombin.
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -!- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14
CC AA repeats.
CC -!- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -!- PTM: Sulfation is required for efficient thrombin cleavage and
CC activation and for full procoagulant activity (By similarity).
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81440; AAA30512.1; -;
DR EMBL; M81441; AAA30513.1; -;
DR PIR; A42580; KFB05.
DR HSSP; P12259; 1CZT.
DR InterPro; IPR001117; Cu-oxidases.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58.C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASEL; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat.
FT SIGNAL. 1 28 POTENTIAL.
FT CHAIN 29 2211 COAGULATION FACTOR V.
FT CHAIN 29 741 COAGULATION FACTOR HEAVY CHAIN (BY
FT SIMILARITY).
FT PROPEP 742 1564 ACTIVATION PEPTIDE (CONNECTING REGION)
FT (BY SIMILARITY).
FT CHAIN 1565 2211 COAGULATION FACTOR LIGHT CHAIN (BY
FT SIMILARITY).
FT FT

FT DOMAIN 30 327 F5/8 TYPE A 1.
FT DOMAIN 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN 203 327 PLASTOCYANIN-LIKE 2.
FT DOMAIN 348 686 F5/8 TYPE A 2.
FT DOMAIN 348 525 PLASTOCYANIN-LIKE 3.
FT DOMAIN 535 686 PLASTOCYANIN-LIKE 4.
FT DOMAIN 696 1564 B.
FT DOMAIN 1124 1151 2 X 14 AA TANDEM REPEATS.
FT REPEAT 1124 1137 1-1.
FT REPEAT 1138 1151 1-2.
FT DOMAIN 1188 1453 30 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
FT REPEAT 1188 1196 2-1.
FT REPEAT 1197 1205 2-2.
FT REPEAT 1206 1214 2-3.
FT REPEAT 1215 1223 2-4.
FT REPEAT 1224 1232 2-5.
FT REPEAT 1233 1241 2-6.
FT REPEAT 1242 1250 2-7.
FT REPEAT 1251 1259 2-8.
FT REPEAT 1260 1268 2-9.
FT REPEAT 1269 1277 2-10.
FT REPEAT 1278 1286 2-11.
FT REPEAT 1287 1295 2-12.
FT REPEAT 1296 1304 2-13.
FT REPEAT 1305 1313 2-14.
FT REPEAT 1314 1322 2-15.
FT REPEAT 1323 1331 2-16.
FT REPEAT 1332 1340 2-17.
FT REPEAT 1341 1349 2-18.
FT REPEAT 1350 1358 2-19.
FT REPEAT 1359 1367 2-20.
FT REPEAT 1368 1376 2-21.
FT REPEAT 1377 1385 2-22.
FT REPEAT 1386 1394 2-23.
FT REPEAT 1395 1403 2-24.
FT REPEAT 1404 1412 2-25.
FT REPEAT 1413 1421 2-26.
FT REPEAT 1422 1430 2-27.
FT REPEAT 1431 1439 2-28.
FT REPEAT 1440 1444 2-29 (PARTIAL).
FT REPEAT 1445 1453 2-30.
FT DOMAIN 1569 1738 F5/8 TYPE A 3.
FT DOMAIN 1569 1738 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1894 2048 F5/8 TYPE C 1.
FT DOMAIN 2053 2208 F5/8 TYPE C 2.
FT SITE 741 742 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1564 1565 CLEAVAGE (BY THROMBIN).
FT DISULFID 167 193
FT DISULFID 248 329
FT DISULFID 499 525
FT DISULFID 607 688
FT DISULFID 1712 1738 PROBABLE.
FT DISULFID 1894 2048 BY SIMILARITY.
FT DISULFID 2053 2208 SULFATION (POTENTIAL).
FT MOD_RES 697 697 SULFATION (POTENTIAL).
FT MOD_RES 701 701 SULFATION (POTENTIAL).
FT MOD_RES 730 730 SULFATION (POTENTIAL).
FT MOD_RES 1513 1513 SULFATION (POTENTIAL).
FT MOD_RES 1529 1529 SULFATION (POTENTIAL).
FT MOD_RES 1537 1537 SULFATION (POTENTIAL).
FT MOD_RES 1541 1541 SULFATION (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .)
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .)
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .)
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .)
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .)
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .)
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .)
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .)
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .)

FT CARBOHYD 774 774 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 780 780 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 902 902 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 952 952 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 964 964 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1044 1044 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1053 1053 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1062 1062 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1071 1071 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1078 1078 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1094 1094 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1451 1451 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1490 1490 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1550 1550 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1690 1690 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1839 1839 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 1997 1997 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT CARBOHYD 2196 2196 N-LINKED (GLNAC. . .) (POTENTIAL) .
FT VARIANT 587 592 NFTLPA -> T (IN VARIANT 2) .
SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B739667C45 CRC64;

Query Match 22.1%; Score 2747; DB 1; Length 2211;
Best Local Similarity 30.2%; Pred. No. 9.2e-134;
Matches 765; Conservative 382; Mismatches 833; Indels 550; Gaps 80;

QY 22 RYVGLGAVLSWDYMQSDLGELPVDARFPFRVPEKSPFFNTSVVYKXTLTFVEFTDHELENIA 81
DQ 32 RQFYVAAQSIWMYR-----PESTHL-----SSKPFETS--FKKIVVREY-EAYFOKE 76

QY 82 KRPAPPWGLGPTIQAEVYDVTIVITLKNMASHVPSLHGVSVYKASBGAEYDDQTSQRE 141
DQ 77 KPOSRTSGLLGPTLYAEVGDIMKVKHAKHAKPLSLHAQGIKYSKPSBGASYSDBHTLPMW 136

QY 142 KEDDKVFPFGSHYVYVQVLKENGPMASDPLCLTYSYLSHVDLVKQNSGLIGALLVCREG 201
DQ 137 KMDDAVAPGOEYIYEWIISHSGPTDHPCLTHIYYSYVNLVEDNSGLIGLLICKKS 196

QY 202 SLAKETQTL--HKFILLFAVPEGKSWHSETKNSLMQDRDAASARAPKMTHTVNGYVNR 259
DQ 197 TLTEGDTQRMFEKQVLMFAVPEDESKSWNQTS-----LMYTVNGVNG 240

QY 260 SLFGLIGCHKSVYVHVGITGTPPEVHSIFLEGHTPLVRNHRQASLEISPIELTIAQTLL 319
DQ 241 TMDPIIVCAHDHLSHGLIGSGGPELFSIHFNQVLEQNHHKSLATLVSATSTTANMTV 300

QY 320 MDLQGLLSCHISSHQHDGMEAYKVVDSCPEEPQLRMKNNEEAEDYDDDLTDSMDVWF 379
DQ 301 SPEGRWTIASLIPRHFOAGQAVIDIKNCAKKTENPKK-----LTRDQ----- 343

QY 380 DDNSPSFTQIRSVAKHPTWHTYIAABEEDWDYAPLVLPADDDRSYKSOYLNNGPQIRG 439
DQ 344 -----RHKIKRWYFIYFAAEVWDYAPIIPANMKKYSRLHLDNFSNRIG 388

QY 440 KKYKVRFMAYTDETEKTR--ZAIQESGTLGPLYGEVGDTLILIFKNOASRPNYIYPH 497
DQ 389 KHYKVVYQYQDDSTKLELDSSEBGDLGILPIIRAQVEDTLKIVFKWASRSYIYPH 448

QY 498 GIT-----DVRPLYSRRLPGVKHLKDFILPEGEIYKWTYVEDGPTKSDPRCL 548
DQ 449 GVTFSDYDNEVNSSTSGSNMIRAVR-----PGETVYKWNILESDEPTENDAQCL 500

QY 549 TRYSSFVNMERDLASGLIGPLIICYKESVDQGNQIMSDKRVNILLFSVPDENRSWYLTE 608
DQ 501 TRYENVDITRLASGLIGELLICKSRLDRGIQRAADIEQQAVFAVFDENKSWIIE 560

QY 609 NIORFLPNFAGVQLEDPEFOASIMH-----SINGVYFDSLQ--LSVCLHEVAVWYILSIG 662
DQ 561 NIYKFCENKVRDDPKFVESIMENFNILPALINGVVPESIPILGCFDDTVQWHFCSVG 620

QY 663 AQTDLSVFPFGYTKHKMVEYDTLTFPFGSETVFMSEMENPGLWILGCHNSDFRNRGMT 722
DQ 621 TQNDILTIHPTGHSFYGKEHEDTLTLFPMQGESVITVMDNVGTMWLTMTWNSNPRSKLR 680

QY 723 ALLKVSSCDKNTGDYVEDSYEDI-----SAYLLSKNNA----- 755
DQ 681 LFRFDAKCIKRNDD--DSYELIYBPSSGTAMTKIHDSSLEEDENADADYQDELALI 737

QY 756 IEPFRSQNSRHPSTROKOFNATTP-ENDIKTDPFWFAHRTPEKPIQNVSSDILMLLR 814
DQ 738 LGLRSFRNSSL--NOEKDELNLTALALEKDSFI PP-SANRS-----LDSNSS----- 783

QY 815 QSPTFEGLSLQCAKYETFPFDDPSGA-----IDSNNSLSEMTHPFQPLHSG----- 864
DQ 784 RSHVRLAKNFABSLKTLHLLEAPAGSPLEBAGLDKNSALNP-----PMAHSSPSYE 838

QY 865 ----DMVFTPEGLOI-----RLNEKLTGTTAATTELKCLDFKVSSTNNL 904
DQ 839 DPREDHPLSDVIGVSLPPGTGFKNKPAPKQFQVGRGQAAKHKSQTFPFAKTRTEL 898

QY 905 ISTIPSDNLAAGTDYSS--LQPPMPVHYDQLDITTLFGKSS-----PLTESGPP 954
DQ 899 -----SQDNSSSRMGP-----WEDIPSDLLLLQKDPYKILNGEWHLYSEKGS 942

QY 955 LSLSENNDKSLLEGLMNSQES--SWGKNVSSSTSGRLFKGKRAHGPAALLT----- 1004
DQ 943 YELIQDANENKTVK-LPNSPONDSTWGEN-PFKNS-----HGKGGHPTFLVTRKPLQ 997

QY 1005 ----KONALFKVSIILKTKNTKNSATNRKTHIDPSILLIENSPSVWQ-----NILES 1054
DQ 998 DRQDRNSRLKEGLPLIR-----TRKKKEKPAYHVPLSPRSPHPLRGEVNASFS 1048

QY 1055 DREFKVTPLIHDRLMDKNATALRLNEMSNKTTSSKNMEMVQKKEGPIPPDAQNPDMS 1114
DQ 1049 DRHN-----HSLILHASNETSLSID--LNQTFPSMNLSLA-----ASLFD-- 1087

QY 1115 FFKMLFLPESAARWIOETHGKNSLN--SGGPPSPKQIV-SLGPEKSVGEGNPLSEKNKVV 1171
DQ 1088 -----HDQSPNTTISQTSPPDLYPTVSPERHYQ----- 1117

QY 1172 GKGETFKDVLGEMVFPSPSRNLFNLNLDNHNTHNNOEKKTOEBIEKBTILQENVLP 1231
DQ 1118 -----IFP-----IQSDPHTSTTAPNSRSP 1139

QY 1232 QIHTVTGTFKNFKNLFLLSSTRONVEGSYD--GAYAPVLQDFRSL----- 1273
DQ 1140 PHTSTAPNSRSP-----PTQSPQIPNYDLRNALPT'DVSOIFPSLELEVWQATSLDLS 1194

QY 1274 NDSVTRTKKHTAFKKGEEENLEGIGNOTKQIVEKYACTTRISNTSQONFVTRSKEA 1333
DQ 1195 QPSISPDLCQMALSPDPGQESLSPLGQ-----TSLSPDLSQESLSPLGQTA 1242

QY 1334 LKQPLPLEET--ELEKRIIVDDTSTQWSKNKHLTP-----STLTQIDYNEKEGKITQ 1386
DQ 1243 LSPD--PSQESLSPLGQATLSPDPS-----QESLSPDLGQATLSPDPGQESLSPLGQ 1294

QY 1387 SPLSDCLTR--SHSIPQANRSPPLIAKVSSPPSRPIVTLRVLFQDNSSHLPAASRYKK 1443
DQ 1295 TSLSPDLSQESLSPLGQATLSPDP-----SQESLSPLGQATLSPD-----PSQESLSPL 1344

QY 1444 DSGVOESSHFLOCAKNNLSLAILTEMTGDQREYGLSGTSSATNSVYKKVNTVLPKPD 1503
DQ 1345 DLGQTSLSPLD-QESLSPLGQATLSPDPQESL----- 1381

QY 1504 LPKTS-----GKVELLPKHVIYQKDLFPETENGSPGHLDIVEGSLLOQTEGAIKWNEAN 1558
DQ 1382 LGQTSLSPLDQESLSPL-----DLGQATLSP-----DLSQESLSPLD----- 1418

QY 1559 RGKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEEWSQESKPEKTAFFKKDTI- 1617
DQ 1419 --GQTP--SPDLSLESJ-----SPDLSQDLDKQTSR 1446

QY 1618 -LSLN-----ACESNHAIAINEGQKPEIEVWAKQGRTERLCSQ-----NPPVLKSH 1665
DQ 1447 PLDILQTSHTSSOSLPLPEFGQIFPNADICQMPSPDPSTLNNTFFIPEENPLVVG 1506

QY 1666 QRE-----ITRTILQSDQEE-----IDYDDTISVENKEDFDIYDEDENQSP----- 1707

```
Db 1507 SRDDGDYIEIPROKESSESDYGBFEFVAYNDPVQTDLR-----DINSRNPDNIAA 1560
Qy 1708 ---RSFOKTHFYFAVERLWDYGMSSSPHVLNRAGSGS---VPQ---PKVVVFQBPFT 1758
Dd 1561 WYLSNTGNRYYYIAAEIISWDYS-----KFVQSDVDYVPEDTVKKVFRKYL 1611
Qy 1759 DGSFTQPLRYGELNHEHLGLGPIYRAEVEDNIMVTFRNOASRPYSFYSLSIYE----- 1812
Dd 1612 DSTFKLPQSEYEHLGILGEVIRAEVDDVLTQVRFKNLARGPISLHAHGLSYEKSSEK 1671
Qy 1813 --EDQOAGPRKNVKNENETKTYPKVQHMAPTKDEFDCXAWAYFSDVDLEKDVHSL 1870
Dd 1672 TYEDDSPEWFKEDNAIQNKTYVYVWHATRSRGPENPCSGACAWAYSAVNPCKDIHSL 1731
Qy 1871 IGPLLVCHTNILNPAHQVQVTPQZPALFTFTDFETKSWYFTENMERNCRAPCNIOMEDPT 1930
Dd 1732 IGPLLLCKGGLDKETNPMDVMDREFVLMVFDEKKSWIYDKKPTRSWRRASS-----E 1785
Qy 1931 FKENRFHAIINGYIMDTLPGLVMAQDQIRWYLLSMGSMENIHSHPGSHVPTVRKKEBY 1990
Dd 1786 VNSHEFHAIINGMIYX-LPGLRMVYQEWVRHLNLNGLGSRDIEHVHFGQTLLENGTQOH 1844
Qy 1991 KVALYNLYPGVPETVEMLPKAGIWEVCLIGEHLHAGNSTLFLVYSNKKOQTPPGMASGH 2050
Dd 1845 QLGVMPLLPKSPKILEMKASKPGWMLLDTVEGEIQRAGWQTPFLIVDRECKMPMLSTGL 1904
Qy 2051 TRDFOITASGOYGQWAPKLARLHYSYNAM-----STK-EPFSKIKVLLAPMI-HGIK 2104
Dd 1905 IADSQIASEFEGWYEPKRLNNGGSSYNAAEKLSTEFPEPWIQVDMQKEVLLTGIG 1964
Qy 2105 TOGARQKSSSIYISOFIIMYSIDGKKWQTYRGNSTGLMVFPGNVDSGIGKHINFPPII 2164
Dd 1965 TOGAKHLKPYTYTEFCVAYSLDRKNWRIFKNSIRNWMYFGNSDASTIKENQIDPFW 2024
Qy 2165 ARYIRLHTHYSIRTLRMELMGDLNCSMPLGMESKAISDAQTASSYFTNMPAT-WS 2223
Dd 2025 ARYIRISPTGYSNKPLALELQGCVEVNGCSTPLGMESGKIENKQITASSFKKSWNGWYE 2084
Qy 2224 PSKARLHLQGRNAPVQNNPKELQVDFQKTMKVTGVTQGVKSLLTSMYVKBFLSS 2283
Dd 2085 PFLARLNAGRRVNAQKANNNNQWLQDLKKIKITAITVQGCCKLSSEMTVKSYTHY 2144
Qy 2284 SQDGHQWILFFONGKV--KVFGNQDSFTPVVNSLDPPLTLRYLRIHPQSWVHQIALRME 2341
Dd 2145 SDQGTDWKPYREKSMWDKIFPGNNNVGRHVKNFNPPIISFIRIIPKTNQSIARLE 2204
Qy 2342 VLGCEAQDLY 2351
Dd 2205 IFGC---DMY 2211

RESULT 7
CERU HUMAN
ID CERU HUMAN STANDARD; PRT: 1065 AA.
AC P00450; Q14063;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Hmo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86259737; PubMed=2873574;
RA Koschinsky M.L., Funk W.D., van Oost B.A., McGillivray R.T.A.;
RT "Complete cDNA sequence of human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5086-5090(1986).
RN [2]
RP SEQUENCE OF 1-1006 FROM N.A.
```

```
RX MEDLINE=95217183; PubMed=7702601;
RA Daimon M., Yamatani K., Igarashi M., Fukase N., Kawanami T.,
RA Kato T., Tomimaga M., Sasaki H.;
RT "Fine structure of human ceruloplasmin gene.";
RL Biochem. Biophys. Res. Commun. 208:1028-1035(1995).
RN [3]
RP SEQUENCE OF 1-40; 549-599; 784-829 AND 919-952 FROM N.A.
RX MEDLINE=86275241; PubMed=3755405;
RA Mercer J.F.B., Grimes A.;
RT "Isolation of a human ceruloplasmin cDNA clone that includes the N-
RT terminal leader sequence.";
RL FEBS Lett. 203:185-190(1986).
RN [4]
RP SEQUENCE OF 218-1065 FROM N.A.
RX MEDLINE=86205876; PubMed=3486416;
RA Yang F., Naylor S.L., Lum J.B., Cutshaw S., McCombs J.L.,
RA Naberhaus K.H., McGill J.R., Adrian G.S., Moore C.M., Barnett D.R.,
RA Bowman B.H.;
RT "Characterization, mapping, and expression of the human ceruloplasmin
RT gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3257-3261(1986).
RN [5]
RP SEQUENCE OF 20-1065.
RX MEDLINE=84119493; PubMed=6582496;
RA Takahashi N., Ortel T.L., Putnam F.W.;
RT "Single-chain structure of human ceruloplasmin: the complete amino
RT acid sequence of the whole molecule.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
RN [6]
RP SEQUENCE OF 158-333; 518-724 AND 858-1065.
RX MEDLINE=83117800; PubMed=6571985;
RA Takahashi N., Bauman R.A., Ortel T.L., Dwulet F.E., Wang C.-C.,
RA Putnam F.W.;
RT "Internal triplication in the structure of human ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
RN [7]
RP SEQUENCE OF 501-905.
RX MEDLINE=81199407; PubMed=6940148;
RA Dwulet F.E., Putnam F.W.;
RT "Complete amino acid sequence of a 50,000-dalton fragment of human
RT ceruloplasmin.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
RN [8]
RP SEQUENCE OF 907-1065.
RX MEDLINE=80137543; PubMed=6987229;
RA Kingston I.B., Kingston B.L., Putnam F.W.;
RT "Primary structure of a histidine-rich proteolytic fragment of human
RT ceruloplasmin. II. Amino acid sequence of the cyanogen bromide
RT peptides.";
RL J. Biol. Chem. 255:2886-2896(1980).
RN [9]
RP SEQUENCE OF 1007-1061 FROM N.A.
RX MEDLINE=90285218; PubMed=2355023;
RA Yang F.Y., Friedrichs W.E., Cupples R.L., Banifacio M.J.,
RA Sanford J.A., Horton W.A., Cupples B.H.;
RT "Human ceruloplasmin. Tissue-specific expression of transcripts
RT produced by alternative splicing.";
RL J. Biol. Chem. 265:10780-10785(1990).
RN [11]
RP REVIEW.
RX MEDLINE=22049919; PubMed=12055353;
RA Hellman N.E., Gitlin J.D.;
RT "Ceruloplasmin metabolism and function.";
RL Annu. Rev. Nutr. 22:439-458(2002).
RN [12]
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
```



```

Db 444 EHLGILGPVIAEWGDTIRVTFNKGAYPLSIEPIGVRFNKNNEGTVYSNYPQSRSP 503
QY 512 KGVELKDPILPGEIFKXKWTYVDEGPTKSPRCLTRYSSFYNNKRDASLGLPPLL 571
Db 504 PSASH-----VAPETFTVETVTFKVEGPTNADPVCCLAKMYSAVDFTKDIFTGLGPMK 558
QY 572 ICYKESVDORGQIMSDKRNVLFSVFDENRSWLTENIORFLPNPAGVQLEDPPEQASN 631
Db 559 ICKKGSILHANGROKVDKRFYLPFTTFDENESILLLEDNIEMFTTAPDQVDKEDDFQBN 618
QY 632 IMESINGYVDSIQ-LSVCLHEVAYWYLLSIGAQTDFLSVFFSGYTFKHKMYVEDTLTLF 690
Db 619 KMFENMGFMYGNQPGTLMCKGDSVWVWYLFSAAGNEADVHGIFYFSGNTYLWRGERDANLF 678
QY 691 PFSGETVFMSEMPGLWILGCHNSDFRNGMTALLKVVSSCDKNTGYDEYSDYEDISAYLL 750
Db 679 PQTSILEWPDTEGFNVECLTIDHYTGKMKQKTYVNOQ----- 718
QY 751 SKNNAIEPRSFSONSRHPTROKQFNATTIPENDIEKTPWFAHRTMPKIQNVSSDDL 810
Db 719 ----- 718
QY 811 MILLRQSPTHGLSLDQAKYETFSDDPSGAIDSNNSLSEMTFRPOLHSHGDMVFTF 870
Db 719 --PROSE-----DSTFYLGERTYY----- 735
QY 871 ESGQLRLNEKLTGTTAATLKKLDFKVSSTSNLLISTIPSDNIAAGTDNTSSLGPPSPV 930
Db 736 -----IAAVEV----- 742
QY 931 HYDSQDHTTFGKKSPFTESGGPLSLEENDSKILLESGLMNSQBSWGKVVSTESGR 990
Db 743 ----- 742
QY 991 LFKGKAHGPAALLTKDNALPKVSIILLKNTKNSNATNRKTHIDGPSLLIENSPSVWQ 1050
Db 743 -----KD- 744
QY 1051 ILESDETFKKTPLIHDRYLMKDNATALNHNKNTSSKNVWQKKEGPIPPDAQ 1110
Db 745 ----- 744
QY 1111 PMSPFKMLFLPESARWIQRTGKNSLNSGQSPKQLVSLGPEKSVGQNFLEKKNV 1170
Db 745 ----- 744
QY 1171 VGKGFTKDVGLKXWVFPSSRNPLFLNLDNLHENNTNHOEKIKQEEIEKKTLIQENVV 1230
Db 745 ----- 744
QY 1231 POIHTVTGTFKNFMKNLFLLLSTRQNEGSYDGAYAPVLQDFRSLNDSTNRKHTAHFSKK 1290
Db 745 -----YSP----- 747
QY 1291 GREENLEGNOTKOIVEKYACTTRISPNTSQNFVQSRKALKQFRLPUBETLEKRI 1350
Db 748 -----QR----- 749
QY 1351 IVDDTSTOWSKNMKELTPTSLTQIDYNEKEKAITQSPLSDCLTSHSIPQANRSPPLIA 1410
Db 750 -----EWEKELHL----- 758
QY 1411 KVSSPFSIRPIYLRVLFDQNSSHLPAASRYKKDSGVQESSHFLQAKKNLSLAITLIE 1470
Db 759 -----GEQVSNAPL----- 768
QY 1471 MTGDQREVGSLGTSATNSVYKVENTVLPKDLPKTSKVELLPKVHIYQKDLFPPTETS 1530
Db 769 ----- 768
QY 1531 NGSPGHLDI-VEGSLLOQTEGALKNEANRPGKVPFLRVATSSAKTPSKLLDPLAWDNHY 1590

```

```

Db 769 ----- 768
QY 1591 GTQIPKEWKSORKSPEKTAFFKKKOTILSNACESHATAINEGQMKPEIEVTWAKOGR 1650
Db 769 ----- 768
QY 1651 TERLCSQNPVLKRGHOREITRITLQSDQBEIDYDDTISYEMKKEPDIYDEDENQSPRSF 1710
Db 769 -----DKGE----- 772
QY 1711 OKKTRHYFIAAVERLMDYGMSSSPHVLNRNRAQSGSVPOFKVVFQFTDGSFTQPLYRGE 1770
Db 773 -----FYIGS-----KXKVVYRCYTDSTFRVPVERKA 800
QY 1771 LNEHLGLLCPYRAEVEDNIMVTFRNQASRPYSFYSSLSISYEDORQGAEPKRNFKV--- 1827
Db 801 EEEHLGILGPQHADVGDKVKIIFKSMATRPYSIHA-----HGVTSSSTVTPTL 850
QY 1828 PNETKTYFMKVOHMAPTKDFDCKAWAYPSVDLEKDVHSGLIGPLLCHTNTLNPAGH 1887
Db 851 PGETITVYWKIPERGAGTESACIPWAYYSTVDQVKDLYSLIGPLIVCRPPLKVFNP 910
QY 1888 RQVTVOEAFALFTTIFDETKSWYFTENMERNCZAPCNIOMEDPTFKENYRFAHNGYIMDT 1947
Db 911 RRKL--EPALLFLVFDENESWLDNINIKTYSDDHPEKWKDDREFIESNKNWHAINGMFCN 968
QY 1948 LFLGVYAQDQORTWLLSMGNSNENIHSFSGHVFTVRKKEBYKMYLNLYPGVFTETM 2007
Db 969 LQGLTMHVGDENVYWMGMNEIDLHTVHFHGHFSQYKRGVYSSDVFDFPQTYQTLEM 1028
QY 2008 LPSKAGIWRVECLIGEHLHAGMSTLFLVYSNK 2039
Db 1029 PRPTGIMLLHCHVTDHIIHAGMETTYTVLQNE 1060

RESULT 8
CERU RAT
ID CERU RAT STANDARD; PRT; 1059 AA.
AC P1355; Q64719;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (ferroxidase).
GN Cp.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver, and Lung;
RX MEDLINE=90237081; PubMed=2332446;
RA Fleming R.E., Gitlin J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
RT tissue-specific gene expression during development.";
RL J. Biol. Chem. 265:7701-7707(1990).
RN [2]
RP SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87137545; PubMed=3818625;
RA Aldred A.R., Grimes A., Schreiber G., Mercer J.F.B.;
RT "Rat ceruloplasmin. Molecular cloning and gene expression in liver,
RT choroid plexus, yolk sac, placenta, and testis.";
RL J. Biol. Chem. 262:2875-2878(1987).
CC -!- FUNCTION: Glycoprotein is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- FUNCTION: May also play a role in fetal lung development or
CC pulmonary antioxidant defense.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers

```



```
QY 1057 EFKKVTELLHDMMLNDKRNATRLNHNMSKNTSSKNMEMVOQKEGPPDPDAQNPDMSEFF 1116
Db 733 ----- 732
QY 1117 KMLFLPESARWQTHGKNSLNSGGPSPKQVLSVIGPEKSVGGNLFSEKKNVVVGKEF 1176
Db 733 ----- 732
QY 1177 TKDVGLEKMPSPSRNLFLTKLDNLHNNHNNHNOEKKIOEEIEKXETLIQENWVLPQIHV 1236
Db 733 ----- 732
QY 1237 TGTQKFMKMLFLLSTRQNVESYDCAVAPVLDQFSLNDSNTRTKHTAHFKKGEENL 1296
Db 733 ----- 741
QY 1297 ECLGNTQKQIVEKVACTTRISPTNSQNFVTRQSKRALKQFRLPLETELEKRIIVDDTS 1356
Db 742 ----- 742
QY 1357 TQSKMKKHLPTSLTQIDYNEKEKGAITQSPCLSTRSHSIPQANRSPPIAKVSSFP 1416
Db 743 RDEWELFHL----- 752
QY 1417 SIRPIYLRVLFQDNSSHLPAASYEKKDSGVQESSHPLOGAKKNLSLAILTEMTGQOR 1476
Db 753 ----- 762
QY 1477 EVGSLGTSATNSVTKYKVENTVLPKPDLPKTSKGVKELLPKVHIQKLPFTTNSGSPQH 1536
Db 763 ----- 762
QY 1537 LDLVEGSLQGTGGAIKNEANRPGVPLRVATESSAKTSKLLDPLANNHYGTQIPK 1596
Db 763 ----- 762
QY 1597 ZEWKSQKSPEKAPKKDITLSLNACSNHAIAINNEGQNKPEIEVTWAKGRTERLCS 1656
Db 763 ----- 762
QY 1657 QNPPVLKRQREITRTTLQSDQEHYDDTISVEMKKEFDIYDEQSPRSFQKTRH 1716
Db 766 ----- 766
QY 1717 VPIAVERLWDYGMSSPHVLNRAQSGSVOPKVVFOEFTDGSFTOPLYRGELNEHLG 1776
Db 767 FFIGS-----KYKKVYREFTDSTFREQVKERABEEHLG 800
QY 1777 LAGPVIRAEVDNIMVTFRNOASRPYSFYSLLSYEEDQOQAEPRKQVK--FNETKT 1833
Db 801 MLGPIHLHADVGAQKVVFKNMATRYSIHA-----HGVKTKSSVAPTLGCEVKT 850
QY 1834 FYKVOHHMAYTKDEFDCKAWAYFSDVLEKDVHSLGLGLPLVC---HTNTLNPAHGRQV 1890
Db 851 YLWQIPERSGAGTEDSPCIPIWAYSVTVDRVKOLYSLGLPLIVCRKSVYKVENPK----- 905
QY 1891 TVQEFALFTIPEETKSVFTNEMERNCRAPCNIOHEDPTFKENFRFAINQYINDTLP 1950
Db 906 KXKEFSLLFLVDENESWYLDNINTPDHPKDKNDNEFTESKKNHAINQKMGFNQ 965
QY 1951 LVMAQDQRIYLLSGNSNENHSHFSQHVFTVRKKEBYKVALNLYPGVFTVEMLPS 2010
Db 966 LTHVGVDEWVWVWANGNIDILHTVHFHSHSQYKHRGIHSSDVFDFPFGTQITLMEFEQ 1025
QY 2011 KAGIWEVECLIGELHAGMSTLFLVYSNK 2039
Db 1026 TSGTWLHCHVTDHAGWVTVTVLPNQ 1054
PRT; 1062 AA.
```

RESULT 9
CERU_MOUSE
ID_CERU_MOUSE

```
AC 061147;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-NOV-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RA Klomp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=96294736; PubMed=8690795;
RA Klomp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;
RL "Ceruloplasmin gene expression in the murine central nervous system.";
RT J. Clin. Invest. 98:207-215(1996).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers
CC known as type 1 or blue, type 2 or normal, and type 3 or coupled
CC binuclear.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues, including liver,
CC eye and brain.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49430; AAB07996.1; -.
DR HSP; P0450; 1KCW.
DR MGD; MGI:88476; Co.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR InterPro; IPR008972; Cupredoxin.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR Transport; Ion transport; Copper transport; Oxidoreductase; Plasma;
DR Metal-binding; Copper; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 1062
FT CERULOPLASMIN.
FT DOMAIN 20 356
FT F5/8 TYPE A 1.
FT DOMAIN 20 199
FT PLASTOCYANIN-LIKE 1.
FT DOMAIN 208 356
FT PLASTOCYANIN-LIKE 2.
FT DOMAIN 369 713
FT F5/8 TYPE A 2.
FT DOMAIN 369 555
FT PLASTOCYANIN-LIKE 3.
FT DOMAIN 565 713
FT PLASTOCYANIN-LIKE 4.
FT DOMAIN 725 1057
FT F5/8 TYPE A 3.
FT DOMAIN 725 896
FT PLASTOCYANIN-LIKE 5.
FT DOMAIN 904 1057
FT PLASTOCYANIN-LIKE 6.
FT DISULFID 173 199
FT BY SIMILARITY.
FT DISULFID 275 356
FT BY SIMILARITY.
FT DISULFID 529 555
FT BY SIMILARITY.
FT DISULFID 632 713
FT BY SIMILARITY.
FT DISULFID 870 896
FT BY SIMILARITY.
FT METAL 120 120
FT COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 990 990
FT COPPER (TYPE 1) (BY SIMILARITY).
```

```

FT METAL 993 COPPER (TYPE 2) (BY SIMILARITY) .
FT METAL 995 COPPER (TYPE 3) (BY SIMILARITY) .
FT METAL 1035 COPPER (TYPE 3) (BY SIMILARITY) .
FT METAL 1036 COPPER (TYPE 1) (BY SIMILARITY) .
FT METAL 1037 COPPER (TYPE 3) (BY SIMILARITY) .
FT METAL 1041 COPPER (TYPE 1) (BY SIMILARITY) .
FT METAL 1046 COPPER (TYPE 1) (BY SIMILARITY) .
FT CARBOHYD 138 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 226 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 583 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 583 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 625 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 757 N-LINKED (GLCNAC. .) (POTENTIAL) .
FT CARBOHYD 922 N-LINKED (GLCNAC. .) (POTENTIAL) .
SQ SEQUENCE 1062 AA; 121159 MW; R3F52ED09A238F16 CRC64;

Query Match 10.0%; Score 1242; DB 1; Length 1062;
Best Local Similarity 18.9%; Pred. No. 1.7e-55;
Matches 393; Conservative 196; Mismatches 432; Indels 1054; Gaps 30;

QY 10 FLCLRFCE-----SATERVYLGAVELSWDYMQ--SDLGELPVDARPPRPVKSFPFN 60
DB 3 FLLSTFIYSSSLALARDKHVFIGITEAVWDYASCTEEKKLISVDTEQSNFYLQNGPDR 52
QY 61 TSVVYKTLFVFTDHLFNIARPPWMLGTLGTIOAENVYDVVITLKNMASHPVSLHAV 120
DB 63 IGRKYKALYFEYTDGTFSTKTDKPAWGLFGLPVKAIVEDKYVHLKXNASRIYTFHAH 122
QY 121 GVSVMKASGEAVDDOTSOREDDKVPFGGSHYVUQVLKENGPMASDPLCLTYSLSH 180
DB 123 GVITYKEYGAVPNDTDFQADKVLFGQQVYVYLHA-NPSPGEGDSNCVTRIYHSH 181
QY 181 VDLKDLNGLIGLALVCREGLAKEKTQTL-HKFTLLPAVDEGKSMSETKNSLM--- 236
DB 182 VDAPXDIASGLIGLILCKKGLYKEKERNIDQEFVIMPVSVDENLSWLEDNKTFCSE 241
QY 237 ---QBRDAASAWPKGTWNGVNRSLPGLICCHRKSYVHVIWGTTFEVHSIFLEGH 293
DB 242 PKVDXNEDFQESNRMYINGVTFGSLPGLNCAADRVKWLFGMGNEVDVESAFHQQ 301
QY 294 TPLVRNHRQASLEISPTTLTAQTLMDLGOFLLSCHISSHQHGWAEVYKVYDSCPEEPQ 353
DB 302 ALTSRNYQTDIINLFATLIDAYMAQNGVWMLSCQNLHLKAGLQAFQVQDC----- 356
QY 354 LRKNNNEEADYDDLDTDSEMDVVRDDNSPFIQIRSVAKKHPTWVHYIAAEEEDWD 413
DB 357 ---NKPSKDK-----NIRKGVH-----YIAAEEVIMN 383
QY 414 YAP--LVLPAPDR-----SYKSOYLNNGPQRIGRKYKURFMAYTDETP---KTREAIQH 463
DB 384 YAPSGIDITFEKLTASGSDSGVFFPQATRIGGSYKQAYREYTDGSTNKRKERGPDEE 443
QY 464 ESGILGPLYGEVGTLLLIIFKNQASRPNIYPHGIT---DVRPLYSRRLPKGVKHLXD 519
DB 444 HLGILGPVIAEWGVDITIKVTENKGOHLSIQPMGVSFPAENEGTYG---PPGASSQQA 500
QY 520 FP-IIPGEITFKYKWTVTVEDGTPSDPRCLTRYSSFFVMNERDLASGLIPLLICVKSIV 578
DB 501 AGHVAPKXTFTYEWTVPKMGPTIADPVCLSKMYGAVDPTKIDFTGLIGPMKKICKKGL 560
QY 579 DQGNQIMSDKRNVLFSVFDENRSMYLTENIQRLPNPAGVLEDPPEQASNMHSING 638
DB 561 LADGRKQVOKKFFYLPFTVFDNESLDDNIRMFTHAPDQVDKEDDEPQSNKXMSMG 620
QY 639 YVFDLSQL-SVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKHXVYEDTLTLPFFSGEIV 697
DB 621 FMYGNQSWPHMCLGESIVWYLPISAGNEADVHGIVYFSGNTYLCKGEERDTANLFPKXSLTL 680
QY 698 FMSMENPGWILICHSNDFENRGMTHALLKVSSCDKNTGDIYEDSYDIDISAYLLSKNNAIE 757
DB 681 LAMPDITKGTDFVBCLTDTDHYTGGMKQKYTVNQCOR-----QFBDFTVYL----- 724
QY 758 PRFSQNSRHSFROKQFNATTIPENDIEKTDQWFAHRTMPKIQNVSSSDJLMLLRQSP 817

```

```

DB 725 ----- 724
QY 818 TEHGLSLSOLQAEKYETFDSDPSPGAIIDSNNSLSEMTFRPOLHSGDMVFTPESGLQIR 877
DB 725 ----- 724
QY 878 LNEKLGTTAAETKLKDFKVSSTSNLSTIPSDNLAAGTONTSSLGPPSPMPVHYDSOLD 937
DB 725 ----- 724
QY 938 TTLFOKKSPLTESGGPLSELSEENNDKLLBSGLMNSQSSWGNVSVSTESGRLEPKGRRA 997
DB 725 -----GERT 728
QY 998 HGPALLTKDONALFKVYSISLLKXNTKSNNSATNRKTHIDGPSLLIENSPSVWONILLESOTE 1057
DB 729 Y-----YVDA----- 733
QY 1058 FKKVTPLIHDMRMDKNATALRLNHSNKTSSKNWEMVQKKKGPIPPDAQNPDMSFPFK 1117
DB 734 ----- 733
QY 1118 MLFELPESARWIORTHGKNSLNSGGQSPKQLVSLGPEKSVBEGQNFSEKNKVVVGKGFT 1177
DB 734 ----- 733
QY 1178 KDVGLKEMVFPSSRNLFJLTNLDNLHENNTHNQEKKIQBEIEKKETLIQBNVVLPIQHTVT 1237
DB 734 ----- 733
QY 1238 GTKNFMKNLFLLLSTRQNVESYDGAYAPVLQDPRSLNDSTNRKHTAHFASKKGEENLE 1297
DB 734 -----VEVEND--YSP----- 742
QY 1298 GLGNQTKQIVKAYACTTRISPNTSQONFVTOQRKRALQFRLPLETELEKRIIYDDTIST 1357
DB 743 -----SR 744
QY 1358 QWSKMKHLLTPSTLTQIDYNEKEKGAITQSPSLSDCLTRSHSIPQANRSPLPPIAKVSSPFS 1417
DB 745 AWEKELHHL----- 753
QY 1418 IRPIYLTVLFDNNSHLPAASVYRKKSQVORSHFLQAKKNLSLAILTLEMTGDQRE 1477
DB 754 -----QEQNV----- 759
QY 1478 VCSLGSATNSVYKKVENTVLPKPDLPKTSKVKVLLPKVHIYQKDLPTFTSNSPGHL 1537
DB 760 ----- 759
QY 1538 DLVEGSLQGTGGAIKWNEANRPGKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKE 1597
DB 760 ----- 759
QY 1598 EWSKQESPEKTAFAKKKDTILSLNACENHAIANEGONKPEIEVWAKQGRTERLCSQ 1657
DB 760 ----- 759
QY 1658 NPPVLKRHQREITRITLOSQBEIDYDDTISVEMKKEDEFDIYDEDESNQSPRSFQKTRHY 1717
DB 760 -----NVFLDKBEFFI----- 770
QY 1718 FIAAVERLWDYGMSSSPHVLNRNRAQSGVFPQPKVYVQFETDGSFTQPL-YRGELNEHLG 1776
DB 771 -----GS--KYKKVYVYQFTDSSPREQVKGRAEDHLG 802
QY 1777 ILGPVIRAEVDNIMVTFERNQASRPYSFYSSLSIYSEBQROQAEPRKNV---KNEKTK 1833
DB 803 ILGPPITHANVGDKVWFKNWATREYSIHA-----HGVKTESSTVVTPLGEVAT 852
QY 1834 YPWKQCHHMAPTKEDFCKAWAYFSDVLEKDVHSLIGLPLVC---HTNTINPAHQV 1890
DB 853 YTWQKPERSGAGREDSACIPWAYYSTVDRKDLVSLGLIPLIVCKSYVKVPSFK----- 907

```

QY 1891 TVQEFALPTTIDETKSWYFTENMBENCRAPCIQWEDTPEKXNRFHAINGVINDTLPG 1950
 Db 908 KXCEFLPLLVDESNWILDNIIKITYSHPKVKNKDNDEEFLSNKOWHAINKMFNQLQG 967
 QY 1951 LVNAQDQIRWYLLSGNSNENIHSIFSGHVTFRKKEEYKMAIYNLYPGVETVEMLPS 2010
 Db 968 LHMVKQDVNVMYLMGNEIDLHTVEFHGHSFOYKRGVYSSDVPFLFSGTQVLTLEMFQ 1027
 QY 2011 XAGINEVECLIGELHAGMSTILFLVYSNKQCTPLG 2045
 Db 1028 TFGTDLHLCHVTCHVHAGMAITTVLPEVEQETKSG 1062

RESULT 10
 MFGM_RAT
 ID MFGM_RAT STANDARD; PRT; 427 AA.
 AC P70490;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-B8) (O-acetyl GD3 ganglioside synthase) (AGS) (MFGM).
 DE acetyl GD3 ganglioside synthase (AGS) (MFGM).
 GN MFG-B8 OR AGS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. [1]_TaxID=10116;
 OX NCBI_TaxID=10116;
 RN [1]_SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96374422; PubMed=8780713;
 RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.; "Cloning and expression of cDNA for O-acetylation of GD3 ganglioside";
 RT Biochem. Biophys. Res. Commun. 225:932-938(1996).
 RL CC -!- FUNCTION: May be involved in phospholipid binding. Seems to participate in the O-acetylation of GD3 ganglioside sialic acid.
 CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
 CC -!- TISSUE SPECIFICITY: Spleen, lung, heart, brain and muscle.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D84068; BAA12210.1; -;
 DR PIR; JC4915; JC4915.
 DR HSP; P00740; 1EDM.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR000421; FAS6 C.
 DR InterPro; IPR008979; Gal bind like.
 DR InterPro; IPR006210; IEGF.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00231; FAS6C; 2.
 DR PROSITE; PS00022; EGF 1; 2.
 DR PROSITE; PS01186; EGF 2; 2.
 DR PROSITE; PS00026; EGF 3; 2.
 DR PROSITE; PS01285; FAS6C_1; 2.
 DR PROSITE; PS01286; FAS6C_2; 2.
 DR PROSITE; PS00022; FAS6C_3; 2.
 DR PROSITE; PS00022; FAS6C_3; 2.
 KW signal; Glycoprotein; Repeat; EGF-like domain; Milk.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 427 LACTADHERIN.
 FT DOMAIN 24 51 EGF-LIKE 1.
 FT DOMAIN 54 108 EGF-LIKE 2.

FT DOMAIN 111 267 F5/8 TYPE C 1.
 FT DOMAIN 272 427 F5/8 TYPE C 2.
 FT DISULFID 28 39 BY SIMILARITY.
 FT DISULFID 33 49 BY SIMILARITY.
 FT DISULFID 51 60 BY SIMILARITY.
 FT DISULFID 68 79 BY SIMILARITY.
 FT DISULFID 73 96 BY SIMILARITY.
 FT DISULFID 98 107 BY SIMILARITY.
 FT DISULFID 111 267 BY SIMILARITY.
 FT DISULFID 254 258 BY SIMILARITY.
 FT DISULFID 272 427 BY SIMILARITY.
 FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 427 AA; 47413 MW; EA8C8631F3EB6047 CRC64;
 Query Match 5.3%; Score 663; DB 1; Length 427;
 Best Local Similarity 40.3%; Pred. No. 3.6e-27;
 Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;
 QY 2018 ECLIGELHAGMSTILFLVYSNK-----COTPLGNASGHIRDFOITASQGY-- 2062
 Db 78 KCLVTEDTQRG--DIFTEYICQCPVGYSGIHCELCGS--KLGLEGAIAQDSQISASSVWG 135
 QY 2063 ----CQWAPKARLHYSGSINAW--STKEPSPWIKVDLLAPMIHGIKTOGARQKPFSSLY 2116
 Db 136 FMGLQWGPFLARLYRTGTVNAWTASSYDSKFWIQVDFLRKMRVSGVMTQGSASRAGRY 195
 QY 2117 ISQFIIMYSLGKWKQVGNSTG--LMVFFGVNDSSGKHNIFPIIARYIRLRHPTHYS 2176
 Db 196 LKTFKVAVSLDGRPEFTQDESGTGDKEFMGNQDNNSKINNFETLBRQYIRLIPVSC 255
 QY 2177 INSTLRMLMGCDLNSCSMPGLMESKAISDAQITASSYFT--NMFA--TWSPSKARLHLQG 2233
 Db 256 RCTLRPELLGCELHGCSEPLGLKNLTIPDSQITASSYKTNWLRAGWYPHLGRLDNQG 315
 QY 2234 RSNAMPQVNPKEWLQVDFQKTKVGTGQVKSILTSYVKEFLISSODGHQWILF 2293
 Db 316 KINAWTQASNAKEWLQVDFQKTKVGTGQVKSILTSYVKEFLISSODGHQWILF 375
 QY 2294 FONGKVKVFGQNCDETFVWNSLDPLTRYLRHHPQSVWVHQAIRMEVLGC 2345
 Db 376 BEQGTSKVFGQVLDNNSKHNIFEXPFVARVVRVPLSWHNRIITLRBLIGC 427
 RESULT 11
 MFGM_MOUSE
 ID MFGM_MOUSE STANDARD; PRT; 463 AA.
 AC P21956; P97800;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-B8) (MFGM)
 DE (Sperm surface protein Sp47) (MP47).
 GN MFG-B8.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NCBI_TaxID=10090;
 OX NCBI_TaxID=10090;
 RN [1]_SEQUENCE FROM N.A.; AND SEQUENCE OF 23-35.
 RC TISSUE=Mammary gland;
 RX MEDLINE=91045008; PubMed=2122462;
 RA Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D., Srinivasan J., Parry G.;
 RT "cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences";
 RT to factor VIII-like sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990). [2]
 RN [2]
 RP SEQUENCE OF 23-463 FROM N.A.

RC TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in phospholipid binding. Zona pellucida-binding protein.
CC -1- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -1- TISSUE SPECIFICITY: Mammary epithelial cell surfaces and spermatozoan.
CC -1- DEVELOPMENTAL STAGE: mRNA expression is detectable in mammary tissue from nonpregnant animals and maximal in the lactating gland.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M39337; AAA39534.1; -;
CC EMBL; Y11684; CAA72380.1; -;
CC PIR; A36479; A36479.
CC HSSP; P00740; 1EDM.
CC MGD; MGI:102768; Mfeg8.
CC DR InterPro; IPR001438; EGF_II.
CC DR InterPro; IPR006209; EGF_like.
CC DR InterPro; IPR000421; FAS8_C.
CC DR InterPro; IPR008979; Gal_Bind_like.
CC DR InterPro; IPR006210; IEGF.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00754; F5_F8_Type_C; 2.
CC PRINTS; PR00010; EGFLOOD.
CC SMART; SM00381; EGF; 2.
CC SMART; SM00231; FAS8C; 2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS50026; EGF_3; 2.
CC PROSITE; PS01285; FAS8C_1; 2.
CC PROSITE; PS01285; FAS8C_2; 2.
CC PROSITE; PS50022; FAS8C_3; 2.
CC Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22
FT CHAIN 23 463
FT DOMAIN 24 61
FT DOMAIN 64 108
FT DOMAIN 148 303
FT DOMAIN 308 463
FT SITE 87 89
FT SITE 28 39
FT DISULFID 33 49
FT DISULFID 51 60
FT DISULFID 68 79
FT DISULFID 73 96
FT DISULFID 98 107
FT DISULFID 145 303
FT DISULFID 290 294
FT DISULFID 308 463
FT CARBOHYD 61 61
FT CARBOHYD 266 266
FT CARBOHYD 316 316
FT CARBOHYD 426 426
FT CONFLICT 30 30
FT CONFLICT 35 35
FT CONFLICT 110 147
FT CONFLICT 168 168
FT CONFLICT 196 196
FT CONFLICT 309 309
FT CONFLICT 395 395
FT CONFLICT 463 AA; 51465 MW; D78B6C6E5F8BA724D CRC64;
FT SEQUENCE

Query Match 5.3%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 8.3e-27;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;
QY 2037 SNKQCPPLGMSGHIRDFOITASGQV-----COWAPKLARLHYSGSINAW--STKEPFS 2088
DB 145 ASKCSQTLGMEGGAIDASQISASVYVNGMGLQWGFELARLYRTGIVNAWNASNDKSP 204
QY 2089 WKVYDLAPMIHGIIXTQARQKFSLSYISQFIIMYSLDGKKWQTYRGNSTGTIMVFFGN 2148
DB 205 WIQWLLLRMRVSGVMTQGASRAGAEYLKTFKVAYSLDGRKFEFIQDESGGD-KEFLGN 263
QY 2149 VDSGGIKHINPPIIARVIRLHPHYSRSTLRMELMGCDLNSCMPLGMSKASDAQ 2208
DB 264 LDNNSLKVNMFNTEAQYIRLYPVSVCHRGCTLRFELLCGLGCLLEPLGLKXNTIPDSQ 323
QY 2209 ITASSYFT--NMEFA--TWSPSKARLHLQGRSNARPOVNNPKEWLQVDFQKTMKVTGVTQ 2265
DB 324 MSASSSKYKTNLRAFQGYHLGLDNGKINAWTAQNSAKELWLDLGTQROVGTIITQ 353
QY 2266 GKSLTSTMYKFLISSQDGHQWTLFPQNGKVKVFGNQDSFTPVVNSLDPLTRYL 2325
DB 384 GARDFGHIQYVESYKVAHSDDGQVMTVEEQSSKVFQGNLDNNSHKKNIFKPFPMARYV 443
QY 2326 RIHPQSVVHIOALRMEVLGC 2345
DB 444 RVLPSVWHARITLRLLELGC 463
ID_MFGM_PIG STANDARD; PRT; 409 AA.
AC P79385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-EB) (MFGM) (Sperm surface protein SP47) (PF47).
GN MFGEB.
OC Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: May be involved in phospholipid binding. Zona pellucida-binding protein.
CC -1- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -1- TISSUE SPECIFICITY: Mammary epithelial cell surfaces and spermatozoan.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y11683; CAA72379.1; -;
CC PIR; T11743; T11743.
CC HSSP; P00740; 1EDM.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR000421; FAS8_C.
CC InterPro; IPR008979; Gal_Bind_like.
CC InterPro; IPR006210; IEGF.
CC Pfam; PF00008; EGF; 2.

```

DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR D3 SMART; SM00231; FA58C; 2.
DR D3 PROSITE; PS00022; EGF_1; 2.
DR D3 PROSITE; PS01186; EGF_2; 2.
DR D3 PROSITE; PS00026; EGF_3; 2.
DR D3 PROSITE; PS01285; FA58C_1; 2.
DR D3 PROSITE; PS01286; FA58C_2; 2.
DR D3 PROSITE; PS00022; FA58C_3; 2.
KW Glycoprotein; Repeat: EGF-like domain.
FT DOMAIN 2 41 EGF-like 1.
FT DOMAIN 44 88 EGF-like 2.
FT DOMAIN 91 247 F5/F8 TYPE C 1.
FT DOMAIN 252 409 F5/F8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; E0C07AF80C29927A CRC64;

Query Match 5.2%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 1.6e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 2018 ECLIGELHAGMSTPLVYSNK-----CQPLGMSGHIRDPQTASQY-- 2062
DB 58 EEVEIDDAHSG--DVFTETICPHGYTGHCHEICNAPLGMETGAIAFQISASSMHG 115
QY 2063 ---GQWAPKLARLVSGSINAW--STKEPFPWIKVDLLAPMIIGIKTGARQKPFSSLY 2116
DB 116 FMGLQWAPLARLRHAGLVNATSNYDNRNFWIQVNLRRMRVTGVVQGSASGASBY 175
QY 2117 ISQFIIMSLDGKKQYTKRGNSTGTILMVFNGVDSGIGKHFNFPDITARYLRHPHYS 2176
DB 176 MKTEKVAYSTDGRKKQFIQGAESGDKIFMGNLDSGLKVNLFEPVLEQVYVRLPIICH 235
QY 2177 IRSTIRMEIMGDLNCSMPGLMESKAISDAQITASSYTN---MFAWSPSKARLHLOG 2233
DB 236 RCTLRFLFELGELSGCABFLGKONTIPNKQITASSFYRTWGLSAFWSYFPYFARLDNQG 295
QY 2234 RSNARPOVANKPEMLQVDFQXTMKVTGVTTOGVKSLLTSMYKBEPLISSODGHQWTLF 2293
DB 296 KFNATQNSASEWLQIDLSQRRVVGITQARDFGHIQVAAKYVSDGVSWTBY 355
QY 2294 FONGKV--KVFGQNDSTFPVNSLDPPLRLYRIHPQSWHQIALRMEVLGC 2345
DB 356 RDQGALEGRKIFFGNLNDSNKKMPETPLTRFVRLPVAWENRITLRVELLGC 409

RESULT 13
MFGM_BOVIN
ID MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; P79344; Q27959;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EG)
DE (MGPS7/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN=Holstein; TISSUE=Mammary gland;

```

```

RX MEDLINE=970C8954; PubMed=8656064;
RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
RA Petersen T.E.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-636(1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGPS7/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391(1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE=Testis;
RA Ensslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -!- FUNCTION: Probably associates with phospholipids on the surface of
CC mammary epithelial cells and milk fat globules. Zona pellucida-
CC binding protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q95114-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q95114-2; Sequence=VSP_001398;
CC -!- TISSUE SPECIFICITY: Milk and spermatozoan.
CC -!- PTM: THE 2 O-LINKED GLYCANS CONSTITUTE OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X91895; CAA62997.1; -.
CC EMBL; S80643; AAB35894.2; -.
CC EMBL; Y11719; CAA72406.1; -.
CC PIR; S74211; S74211.
CC HSP; P00740; IIXA.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR000421; FA58C.
CC InterPro; IPR008979; Gal_bind_like.
CC InterPro; IPR006210; IEGF.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00231; FA58C; 2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS00026; EGF_3; 2.
CC PROSITE; PS01285; FA58C_1; 2.
CC PROSITE; PS01286; FA58C_2; 2.
CC PROSITE; PS00022; FA58C_3; 2.

```


Signal: Glycoprotein; Milk; Repeat; EGF-like domain;
KW Alternative splicing.
FT SIGNAL 1 18
FT CHAIN 19 427 LACTADHERIN.
FT DOMAIN 20 59 EGF-LIKE 1.
FT DOMAIN 62 106 EGF-LIKE 2.
FT DOMAIN 109 265 F5/8 TYPE C 1.
FT DOMAIN 170 427 F5/8 TYPE C 2.
FT SITE 65 87 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 24 35 BY SIMILARITY.
FT DISULFID 29 47 BY SIMILARITY.
FT DISULFID 49 58 BY SIMILARITY.
FT DISULFID 66 77 BY SIMILARITY.
FT DISULFID 71 94 BY SIMILARITY.
FT DISULFID 96 105 BY SIMILARITY.
FT DISULFID 109 265 BY SIMILARITY.
FT DISULFID 252 256 BY SIMILARITY.
FT DISULFID 270 427 BY SIMILARITY.
FT CARBOHYD 27 27 O-LINKED (FUC. . .) (IN PAS-6).
FT CARBOHYD 34 34 O-LINKED (FUC. . .) (IN PAS-7).
FT CARBOHYD 59 59 N-LINKED (GLNAC. . .) (HYBRID)
FT CARBOHYD 227 227 N-LINKED (GLNAC. . .) (HIGH MANNOSE)
FT VARSPLIC 169 221 Missing (in isoform Short).
FT CONFLICT 19 19 A -> P (IN REF. 1).
FT CONFLICT 28 28 L -> O (IN REF. 1).
FT SEQUENCE 427 AA; 47411 MW; 4CBEE3A1DC4E24 CRC64;

Query Match 5.1%; Score 635; DB 1; Length 427;
Best Local Similarity 37.9%; Pred. No. 1e-25;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGHLLHAGSTLFLVSNK-----CQPLGASGHIRPQITASQY-- 2062
Db 76 ECQVTDSDRG--DVFIQVTKCPLOYGVHCHETCTSPGMQTGAIDSGIASMHLG 133

QY 2063 ----GOWAPKLARLHYSGSINAWST---KEPFSWIKVDLAPMIITHGKIQGARQKFS 2114
Db 134 FMLQWAPLALHQTGINAWTSYNDKPN--WQVNLMRKWVTVGVVQASRAGSA 191

QY 2115 LIISQFIINSLDGKQWYRGNSTGLTVFFQGVNDSSGKHNIFNPPIARIYRLHPTH 2174
Db 192 EYLKTEPKVAYSTDGRQFQFQVQAGSGDKIFIGNVANSGLKINLPTPLETVQLVLPV 251

QY 2175 YSRSTLRLMELMGCDLNSCMPLGMSKALSDAQITASSYFTN---MPATWSPKARLHL 2231
Db 252 CHRGCTLRPELLGCELANGTEPEGLKNDTIPNKQITASSYFTWGLSAPFSPFYARLDN 311

QY 2232 QGRSNARFQVNNPKWLQVDFQKTMKVTVGVTTQGVKSLTSMYKFEFTISSQDGHQWT 2291
Db 312 QGKFNATWATQNSASEMLQIDLSQKEVTGIITQGARDFGHIQVVAAYRVAGDDGVTT 371

QY 2292 LFFQNG--KVKVQGNQDSFTFVNSLDPLLPRIYRIHPQSWHQAIRMEVLGC 2345
Db 372 EYKDPGASESKIPFGNNDNNSHRKNIFETPFQARFVRIQPVAMHNRTLRVELLGC 427

RESULT 14
MFGM_HUMAN
ID MFGM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EB) (HMFG)
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].
GN MFGEB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.
RP TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213908; PubMed=8639264;
RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
BA46 reveals an RGD cell adhesion sequence presented on an epidermal
growth factor-like domain";
RL DNA Cell Biol. 15:281-286(1996).
[2] SEQUENCE OF 170-387 FROM N.A.
RP TISSUE=Mammary Gland;
RX MEDLINE=91371351; PubMed=1909932;
RA Larocca D., Peterson J.A., Ureia R., Kuniyoshi J., Bistrain A.M.,
Ceriani R.L.;
RT "A Mr 46,000 human milk fat globule protein that is highly expressed
in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
[3] PARTIAL SEQUENCE, AND CHARACTERIZATION.
RP TISSUE=Milk;
RX MEDLINE=98194924; PubMed=9535276;
RA Giuffrida M.G., Cavalletto M., Giunta C., Conti A.,
Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
[4] SEQUENCE OF 269-317, AND IDENTIFICATION OF MEDIN.
RP MEDLINE=99342076; PubMed=10411933;
RX Haeggqvist B., Naeslund J., Sletten K., Westermark G.T., Mucchiano G.,
Tjernberg L.O., Nordstedt C., Engstrom U., Westermark P.;
RT "Medin: an integral fragment of aortic smooth muscle cell-produced
lactadherin forms the most common human amyloid.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
[5] CHARACTERIZATION.
RP MEDLINE=97405885; PubMed=9260929;
RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
(RGD)-dependent cell adhesion";
RL DNA Cell Biol. 16:861-869(1997).
CC -!- FUNCTION: May be involved in phospholipid binding. Binds
specifically to rotavirus and inhibits its replication.
CC -!- FUNCTION: Medin is the main constituent of aortic medial amyloid.
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -!- TISSUE SPECIFICITY: Mammary epithelial cell surfaces and aortic
media. Overexpressed in several carcinomas.
CC -!- PTM: MEDIN HAS A RAGED N-TERMINUS WITH MINOR SPECIES STARTING AT
AMINO ACID 264 AND 273.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL; U58516; AAC50549.1; -
EMBL; S56151; AA319771.1; -
FIR; A47285; A47285.
HSSP; P08709; 1BF9.
Genew; HGNC:7036; MFGEB.
MIM; 602281; -
GO; GO:0007155; P-cell adhesion; TAS.
InterPro; IPR001438; EGF_II.
InterPro; IPR006209; EGF_like.
InterPro; IPR000421; FA56C.
InterPro; IPR008979; Gal_bind_1.

```

DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFELOOD.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00022; FAS8C_3; 2.
KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT SIGNAL 1 23
FT CHAIN 24 387
FT CHAIN 202 387
FT CHAIN 285 317
FT DOMAIN 24 67
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 46 48
FT SITE 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
SQ SEQUENCE 387 AA; 43123 MW; 2EB6571DEC83782D CRC64;

Query Match 4.7%; Score 588; DB 1; Length 387;
Best Local Similarity 37.3%; Pred. No. 2.3e-23;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;

QY 2066 EMLPGKAGIWRVECLIGELHAGMSTFLVYNSKCTPLCMASGHIRDFOITAGS----- 2060
DB 48 DYFSPS-----YTCTCLKG---YAGNHC-----ETKVEPLGMEGNTANSQIASSVRVTF 95
QY 2061 -QYGWAPKIALHLYGSGINAW--SKYEFMSIKVDLLAPZMIHGHKTQCARQKPSLVI 2117
DB 96 LGLQHWVPELALINRAGMVAWTFSSNDNDNPWITQVNLRLRWVTVGTQASRLASHEYL 155
QY 2118 SFIIMYSIDGKKWQYRGNSTCTLMVRFSGVDSSGIKENINPPIIASYVIRLHPHYSI 2177
DB 156 KAFKVAISLNGHEFD-FIHDVKKHKEFVGNMKNNAVHVNLFETPVEAQYVRLYPTSCHT 214
QY 2178 RSLRLMELMGCLINSMSMELGMSKASDAQITASSYF----TNMFATWSPGKARLHLQ 2233
DB 215 ACTLRFELGCELGCANPLGKNNISIPDKQITASSSYKTWGLHLP-SWNPYSYARLDKQ 273
QY 2234 RSNAMPQVNNPKMLOVDFQKTMKVTGTTQVCKSLTMSYVKEPLISSSQDGHQWTLF 2293
DB 274 NFNAVAGSYGNDQMLQVLDLGSKEVTGIIQTQARNFGSVQFVASYKVAYSNDSANWTEY 333
QY 2294 F--ONGKVKVFCGQNDSPFVNVSLDPLLTLYRLTHPQSWHQAIALRMEVLGC 2345
DB 334 QDPRGSSKIFPGNWDNHHKQMLFETPILARYIRLPVAVHNRIRARLELLGC 387

RESULT 15
ID NRP2_HUMAN STANDARD; PRT; 931 AA.
AC Q60462; Q14820; Q14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neupilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2)
GN NRP2 OR VEGF15R2.
OS Homo sapiens (human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE=97470889; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neupilin-2, a novel member of the neupilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III.";
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RC TISSUE=Breast;
RX MEDLINE=98189099; PubMed=9529250;
RA Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT "Neupilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor.";
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen Y., Herzog Y., Neufeld G.;
RT "Neupilin-2 and neupilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neupilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF.";
RL J. Biol. Chem. 275:18040-18045(2000).
CC -!- FUNCTION: High affinity receptor for semaphorins 3C, 3F, VEGF-165
CC and VEGF-145 isoforms of VEGF, and the PlGF-2 isoform of PGF.
CC -!- SUBUNIT: Neupilin-2 probably forms a heteromeric complex with
CC neupilin-1 in order to be a functional semaphorin 3C receptor.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=3;
CC Name=A22;
CC IsoId=O60462-1; Sequence=Displayed;
CC Name=A0;
CC IsoId=O60462-2; Sequence=VSP_004342;
CC Name=A17;
CC IsoId=O60462-3; Sequence=VSP_004341;
CC -!- SIMILARITY: Belongs to the neupilin family.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022859; AAC51788.1; -;
CC EMBL; AF022860; AAC51789.1; -;
CC EMBL; AF016098; AAC12922.1; -;
CC HSSP; PI2259; 1CZT
CC Genew; HGNC:8005; NRP2.
CC MIM; 602070; -;
CC GO; GO:0005624; C:membrane fraction; TAS.
CC GO; GO:0004872; F:receptor activity; TAS.
CC GO; GO:0005021; F:vascular endothelial growth factor receptor. . .; TAS.
CC GO; GO:0007411; P:axon guidance; TAS.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR000421; FAS8C.
CC InterPro; IPR008979; Gal_Bind_like.
CC InterPro; IPR000998; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC Pfam; PF00629; MAM; 1.
CC PRINTS; PR00020; MAMDOMA1N.
CC SMART; SM00042; CUB; 2.

DR SMART; SM00231; FA58C; 2.
DR SMART; SM00137; MAX; 1.
DR PROSITE; PS0118C; CUB; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00060; MAM 2; 1.
KW Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW Alternative splicing
FT SIGNAL 1 20 OR 22 (POTENTIAL).
FT CHAIN 21 931 NEUROFILIN-2.
FT DOMAIN 21 864 EX-TRACELLULAR (POTENTIAL).
FT TRANSMEM 865 889 POTENTIAL.
FT DOMAIN 890 931 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 28 142 CUB 1.
FT DOMAIN 149 267 CUB 2.
FT DOMAIN 277 427 FS/8 TYPE C 1.
FT DOMAIN 434 592 FS/8 TYPE C 2.
FT DOMAIN 642 802 MAM.
FT DOMAIN 671 674 POLY-SER.
FT DISULFID 28 55 BY SIMILARITY.
FT DISULFID 83 105 BY SIMILARITY.
FT DISULFID 149 175 BY SIMILARITY.
FT DISULFID 208 230 BY SIMILARITY.
FT DISULFID 277 427 BY SIMILARITY.
FT DISULFID 434 592 BY SIMILARITY.
FT CARBOHYD 152 152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 809 813 Missing (in isoform A17).
FT VARSPPLIC 809 830 Missing (in isoform A0).
FT CONFLICT 602 602 E->K (IN REP. 1).
FT SSSEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;

Search completed: April 13, 2004, 14:05:08
Job time : 39 secs

Query Match 3.8%; Score 469.5; DB 1; Length 931;
Best Local Similarity 28.3%; Pred. No. 1.2e-16;
Matches 166; Conservative 87; Mismatches 185; Indels 145; Gaps 32;
QY 1847 DEFDCX-ANAYPSDVLDLEK-----DVHSLGIPLLVCHTNL-----NPAHGRQVTOEFA 1896
DB 79 EKHDCKYDFTEIRDGSEADLLGKHCNIAPTIISSGMLYIKFTSDYARQGA--GFS 136
QY 1897 LFTTFDETSWYPTENMERNCRAPCNIOMEDPTFKENYRPHAINGYIMDTLPGLVMAQD 1956
DB 137 LRYEIP-KTGS-----EDCSNFTSP-NGTIESGFPPEKP----- 170
QY 1957 QRIRWYLLSGSNENIHSIHPGSHVFTVRKKEYYKALYNLYPGVFTVEMLPSKAG-- 2013
DB 171 -----ENLDCFTILAKPKMEIILQFL--IFD-LEHDPQLQVGEGD 207
QY 2014 -----IW-----RVECLIGH-----LHAGMST-----LFLV 2035
DB 208 CKYDWDLDWGIPIHVGVLGKYCGTKTPSELRSSTGILSTFFHDMAVAKDGFSAARYLV 267
QY 2036 YSN-----KQOTPLGMASGHIRDPQITASQY--GOWAPKLARLHYSGSINAW-----ST 2083
DB 268 HQELENFQCNVPLGMSGRIANEQISASTYSDGRWTPQOSRLH--GDDNGWTFNLDN 325
QY 2084 KPPPSWKVLLAPMIHGIKTOGA--RQPFSSLYISQFIMYSLDGKKWQTVRGNSTGT 2141
DB 326 KE---XILQVDLRLFTLMTATATQAGISRETONGYVXSYKLEVSTNGEDWVYRHGKNH- 381
QY 2142 LMVFFGVNDSSGIGKHNIFNPILIARYIRLPHETHYSIRSTIEMELMGCDLNS--CSMP2LGM 2199
DB 382 -KVFQANDAEVVKLKHAPLLTRFVRIRPQTHSGIALRLLELFGCRVTDAPCSNMLZM 440
QY 2200 ESKAISDAQITASSYTNMPATMSPSKARLHLOGRSNAW---RQVNNPKEWLOVDFQKT 2256
DB 441 LSGLIADSOISASS--TQEY-LNPSRRL-VSSRS-GWFPRIQAPGGEWLQVDLQTP 495

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 68.5 seconds
(without alignments)
10828.953 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MQIELSTCFLLCLRLFCFSA.....WVHQIALRMVLGCGAQQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9485	76.4	2343	6 O18806	O18806 canis famil
2	9439	76.0	2343	6 O62730	O62730 canis famil
3	6283.5	50.6	2258	11 Q7TN96	Q7TN96 rattus norv
4	4043.5	24.5	1639	13 Q804W6	Q804W6 fugu rubrip
5	2825.5	22.8	2119	13 Q90X47	Q90X47 brachydanio
6	2774.5	22.3	2183	11 O88783	O88783 mus musculu
7	2698	21.6	1802	13 Q804W5	Q804W5 fugu rubrip
8	2511.5	20.2	1377	13 Q804X3	Q804X3 gallus gall
9	2340	18.8	1460	13 Q7SZN0	Q7SZN0 pseudonaja
10	1999.5	16.2	2102	11 Q7TFK2	Q7TFK2 rattus norv
11	1595.5	12.9	355	11 Q8BQ43	Q8BQ43 mus musculu
12	1453	11.7	745	13 Q804X4	Q804X4 gallus gall
13	1389	11.2	1156	11 Q80Y80	Q80Y80 mus musculu
14	1388	11.2	1157	11 Q920Z4	Q920Z4 mus musculu
15	1374	11.1	1157	11 Q920H8	Q920H8 rattus norv
16	1344	10.8	1158	4 Q9BQ57	Q9BQ57 homo sapien

17	1339	10.9	1158	4	Q9C058	Q9C058 homo sapien
18	1328	10.7	1087	13	Q7ZUL2	Q7ZUL2 brachydanio
19	1307	10.5	1104	4	O75180	O75180 homo sapien
20	1298	10.5	1084	11	Q9JL97	Q9JL97 rattus norv
21	1227.5	9.9	1048	6	Q9XT27	Q9XT27 ovis aries
22	1123.5	9.0	847	11	Q8C4S2	Q8C4S2 mus musculu
23	1104	8.9	216	4	Q14286	Q14286 homo sapien
24	932	7.5	626	13	Q90ZT2	Q90ZT2 brachydanio
25	911	7.3	407	13	Q8AY50	Q8AY50 brachydanio
26	743.5	6.0	1142	10	Q8LL91	Q8LL91 chlamydomon
27	700	5.6	503	11	Q8BV37	Q8BV37 mus musculu
28	664	5.3	463	11	Q9R1X9	Q9R1X9 mus musculu
29	662.5	5.3	480	4	Q8N610	Q8N610 homo sapien
30	662.5	5.3	480	4	O43854	O43854 homo sapien
31	661	5.3	426	11	Q9WTS3	Q9WTS3 mus musculu
32	655.5	5.3	470	11	Q8C4U8	Q8C4U8 mus musculu
33	655.5	5.3	480	11	Q8CBF7	Q8CBF7 mus musculu
34	655.5	5.3	480	11	O35474	O35474 mus musculu
35	555	4.5	312	4	Q7Z3D2	Q7Z3D2 homo sapien
36	514.5	4.1	363	6	O77718	O77718 equus cabal
37	488	3.9	335	4	Q9BTL9	Q9BTL9 homo sapien
38	469.5	3.8	901	4	Q9H2E4	Q9H2E4 homo sapien
39	469.5	3.8	901	4	Q9H2D5	Q9H2D5 homo sapien
40	469.5	3.8	906	4	Q9H2D4	Q9H2D4 homo sapien
41	469.5	3.8	906	4	Q9H2E3	Q9H2E3 homo sapien
42	469.5	3.8	931	4	Q7Z3T9	Q7Z3T9 homo sapien
43	466.5	3.8	384	11	Q8C8K0	Q8C8K0 mus musculu
44	462.5	3.7	926	11	Q8QZT7	Q8QZT7 mus musculu
45	450.5	3.6	921	11	Q9QX38	Q9QX38 rattus norv

ALIGNMENTS

RESULT 1

O18806 PRELIMINARY; PRT; 2343 AA.
 AC O18806;
 DT 01-JAN-1998 (TREMREL. 05, Created)
 DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
 DT 01-OCT-2003 (TREMREL. 25, Last annotation update)
 DE Factor VIII.
 GN F8.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxId=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Cameron C., Notley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,
 RA "The canine factor VIII cDNA and 5' flanking sequence."
 RT Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 2 FS/8 TYPE C DOMAINS.
 DR EMBL; AF016234; AAB87412.1; .
 DR HSSP; P00451; 1CFG.
 DR GO; GO:0005507; E: copper ion binding; IEA.
 DR GO; GO:0007155; P: cell adhesion; IEA.
 DR InterPro; IPR001117; Cu-oxidase.
 DR InterPro; IPR008972; Cupredoxin.
 DR InterPro; IPR000421; FA58 C.
 DR InterPro; IPR008979; Gal Bind_like.
 DR Pfam; PF00394; Cu-oxidase; 3.
 DR Pfam; PF00754; F5_F8_type_C; 2.
 DR SMART; SM00231; FA58C; 2.
 DR PROSITE; PS01285; FA58C.1; 2.
 DR PROSITE; PS01286; FA58C.2; 2.
 DR PROSITE; PS00022; FA58C_3; 2.
 DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
 SQ SEQUENCE 2343 AA; 258629 NW; A854FAE571C3B399 CRC64;

Query Match

76.4%; Score 9485; DB 6; Length 2343;

Best Local Similarity 77.1%; Pred. No. 0;
Matches 1816; Conservative 196; Mismatches 320; Indels 24; Gaps 12;

QY 1 MQBELSTCFCLLRFCSATRRYYLGAVELSWDMSD-LGELPVDARPPRPVKPF 59
DB 1 MQVELYTCCLLCLLPFSLATRKYYLGAVELSWD;MOSDLLSALHATDSFSSRVEGSDPL 60
QY 60 NTSVYKTLFVEFTDHLFNIAKPRPPMMGLLGFTIAQEVYDVTIVILKNWASHPVSLHA 119
DB 61 TTSVYKTLFVEFTDHLFNIAKPRPPMMGLLGFTIAQEVYDVTIVILKNWASHPVSLHA 120
QY 120 VGVSYWASGAEEVDDQTSQREKEDDKVFPQGSHTYVQVILKENGPMASPLCTYSYLS 179
DB 121 VGVSYWASGAEEVDDQTSQREKEDDNVFPQGSHTYVQVILKENGPMASPPCLTYSYFS 180
QY 180 HVDLVKDLNSGLIGALLVCBEGSLAKCTQTLKFIPLFAVFDGKSWHSSTKNSLMQDR 239
DB 181 HVDLVKDLNSGLIGALLVCBEGSLAKCTQTLQFVLLFAVFDGKSWHSSTKNSLMQDR 238
QY 240 DAASARAWPMHTVGVNRSPLGLIGCHRSVYVHVGITTEPEVHSIFLEGHTFLVRN 299
DB 239 ---AEAQHELHTNGVNRSLPGLTVCHKRSVYVHVGITTEPEVHSIFLEGHTFLVRN 294
QY 300 HROASLEISPIFETACTALLMDLQFLLSCHISSHODGMEAYKYVDSCEBEPOLRMKN 359
DB 295 HROASLEISPIFETACTALLMDLQFLLPCHIPSHODGMEAYKYVDSCEBEPOLRMKN 354
QY 360 BEAEDYDDDLTDSMDVVRPDDNSPFIQIRSVAKXHPKTHVHYIAAEEDWDYAPVL 419
DB 355 ED-KDYDDGLVSDMDVVRPDDNSPFIQIRSVAKXHPKTHVHYIAAEEDWDYAPVL 413
QY 420 APDRYSKSOVLNNGPORIGRYKXKVPFMAVDTFTKTRBAIQHSGTILGPLYGEGVDT 479
DB 414 TPNDRSKNLYLNNGPORIGRYKXKVPFMAVDTFTKTRBAIQHSGTILGPLYGEGVDT 473
QY 480 LLIIFKQASBPYNIYPHGTCTZPLYSRLPKGVHLKOPPLPGSIFKYKWTVTYEDG 539
DB 474 LLIIFKQASBPYNIYPHGTCTZPLYSRLPKGVHLKOPPLPGSIFKYKWTVTYEDG 533
QY 540 PTKSDPCLTRYSSFFNMERDLASGLTGPLLICYSVDRGNQIMSKDNVILFSVFD 599
DB 534 PTKSDPCLTRYSSFFINERDLASGLTGPLLICYSVDRGNQIMSKDNVILFSVFD 593
QY 600 ENRSWYLTENLORELPNAGVQLSDPFPQASNIMHSINGVYFVDSLOISVCLHEVAYWYIL 659
DB 594 ENRSWYLTENLORELPNAGVQLSDPFPQASNIMHSINGVYFVDSLOISVCLHEVAYWYIL 653
QY 660 SIGAQTDPLSVFSGYTFKXWVYEDTTLTLPFSGETVEMSMENPGIWLGCNHSDFRNR 719
DB 654 SVGAQTDPLSVFSGYTFKXWVYEDTTLTLPFSGETVEMSMENPGIWLGCNHSDFRNR 713
QY 720 GMTALLKVSSCDKNTGYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATT 779
DB 714 GMTALLKVSSCNRIIDDYEDYEDIEDIPTPLNENNVIKPRSFQNSRHPSTROKQFNATT 773
QY 780 IPENDIEKTDPAHRTFPMKIQNVSSDILMLRQSPTHGILSLDQAKYETFSDDP 839
DB 774 TPENDIEKIDQSERQOLKAOQVSSDILMLRQSPTHGILSLDQAKYETFSDDP 831
QY 840 SPGAIDNSNLSSEMTFRPQLHHSQDMVFTPEGLQLRLNEKLTGTAATKLLDKFVSS 899
DB 832 SRGAIERKGPPEVASLRPELRHSEDEFTPEPELQRLNENIGTNTVELKLLDKISS 891
QY 900 TSNNLIS--TIPSONLAAGTNTSSLGPPNPHVYDSQOLDTILFGKXSSPLTESGGLSL 957
DB 892 SSDSLMTSPITPSDKLAATEKTSGLGPPNMSVHFNSHLGTIVFGNNSHLIQSGVPLEL 951
QY 958 SEENNDSKLESGLANQESGWGNVSTSGRLPKRAKAGPALLTKDNALFKVSL 1017
DB 952 SEENNDSKLEAPLANTQESLSRNLVMSNRLFKERIRGPFASLTKDNALFKVSSV 1011
QY 1018 KTNKTSNNSATNRKTHIDGFSLLTENGSPVWQNT-LESDFEFKVTPLIHDRMLMKNAT 1076

DB 1012 KTRRAPVNLATNRKTRVAIPTLLIENSTSVWQDILMERNTETFEKVTSLIHNETFMDRNTT 1071
QY 1077 ALRLNEMSNKTTSSKNMENVQOKKEGPIPDQONPOMSFKMLFLPESARWIOETHCKNS 1136
DB 1072 ALGLNEVSNKTTLSKNVEMAHQKEDPVPURAEENPDLSSEKIFPLPD---WIKTHGNS 1127
QY 1137 LNSGQSPSPQJLSLOPEKSVQGNFLSEKNVVGKFTKDVGLKEMVFPSSRNILFT 1196
DB 1128 LSSBQSPSPQJLSLOPEKSVQGNFLSEB-KVVNGDEFTKDELOE-IFPNKKSIFPA 1185
QY 1197 NLNLENHNTNOKKIQEBIEKKEITLIQENVLPQIHTVGTGNPMKNLFLLSSTRONVE 1256
DB 1186 NLANVOENDYNECKSPBIEKKEITLIQENVLPQIHTVGTGNPMKNLFLLSSTRONVE 1245
QY 1257 GSYDGAVAPVLOPFRSLNDSNTNKKHTAHFKKGBEENLEGLGNOKQIVEXVACTTRI 1316
DB 1246 GLEQPYTPILQOTRSLNDSNPHSEGHIMANFSKIRBEANLEGLGNQINQWVERPPSTTR 1305
QY 1317 SPNTSQNFYQSRKALKQFRLEBETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDY 1376
DB 1306 SSNAS-QHVITQRGKSLKQPRLSQGEIKFERKVIANDTSTQWSKNMYLAQGTLTQIEY 1364
QY 1377 NEKEKGAITOSPISDCLTRSHSIPQANRSELPITAKVSSPFSIRPIYLTIVLPODSSHLP 1436
DB 1365 NEKEKGAITOSPISDCLTRSHSIPQANRSELPITAKVSSPFSIRPIYLTIVLPODSSHLP 1424
QY 1437 AA---SYRKKGDSGVQESSHFLQAKKNLSLAILTLEMTGDQREVGSLSGTSATNSVYK 1492
DB 1425 ASACNTYFRERTSGVQEGSHFLQAKKNLSLAILTLEMTGDQREVGSLSGTSATNSVYK 1484
QY 1493 KVENTVLKPDLPKTSKVELLPKVIYKQDLPPTTSNGSPGHLDLVBESLLQGTGAI 1552
DB 1485 KLENTVLLQGLSETDKVELLSQVHVQDSDSPTKTSNDSPGHLDLIMGKIFLQKQTGPV 1544
QY 1553 KWEANRPKVPFLRVATESSAKTPSKLLDPLANDNEHYCTQIPKEBKSSEKSPKTAFAK 1612
DB 1545 KMNKTSNPGKVPFLKWAATESSEKIPSKLLGVLAWDNEHYCTQIPSEBKSSEKSPKTAFAK 1604
QY 1613 KCDTILSLNACENHAIAAINEGONKPEIETVWAKQRTERLCSQNPVPLKRRQREITRT 1672
DB 1605 RKDTILPLGFCENNDSIAINEGONKPEIETVWAKQRTERLCSQNPVPLKRRQREITRT 1664
QY 1673 TLOSQBEIDYDTISVENMKEDFDYDDEBENOSPRSFOKKTTHYFIAAVERLWDYGMSS 1732
DB 1665 TLOSQBEIDYDTISVENMKEDFDYDDEBENOSPRSFOKKTTHYFIAAVERLWDYGMSS 1724
QY 1733 SPVLNRNAGSGSVPOFKVVFQSTDPQSTPOLYRGELENEHLGLLGPYIRAEVENIMV 1792
DB 1725 SPILLNRNAGSGSVPOFKVVFQSTDPQSTPOLYRGELENEHLGLLGPYIRAEVENIMV 1784
QY 1793 TFEQASRPYSPYSSLSIYEEDQORQGAERPKNFVKPNETKTYFWKQVHMAPTKDFDCK 1852
DB 1785 TFEQASRPYSPYSSLSIYEEDQORQGAERPKNFVKPNETKTYFWKQVHMAPTKDFDCK 1844
QY 1853 AWAYFSDVLEKDVHSLGLPLLVCHTNILNPAHQVQVTVQEFALPTTFDETKSWYFTE 1912
DB 1845 AWAYFSDVLEKDVHSLGLPLLVCHTNILNPAHQVQVTVQEFALPTTFDETKSWYFTE 1904
QY 1913 NMERNCRAPCNIOQVEDPTTKENYRFFHAINGYIMDTLPLGVMAQDQRIWVYLLSMGSENI 1972
DB 1905 NLERNCRAPCNIOQVEDPTTKENYRFFHAINGYIMDTLPLGVMAQDQRIWVYLLSMGSENI 1964
QY 1973 HSIHFSCHVFTVRKSEYKMAALNLYPGVETVEMLPKAGIWRVRECLIGEHLHAGMSTL 2032
DB 1965 HSIHFSCHVFTVRKSEYKMAALNLYPGVETVEMLPKAGIWRVRECLIGEHLHAGMSTL 2024
QY 2033 FLVYSKCCQTPJGASGHIRDFOITASQGVQWAPKALRLHYSGSINAWSTKPPPSWIKV 2092
DB 2025 FLVYSKCCQTPJGASGHIRDFOITASQGVQWAPKALRLHYSGSINAWSTKPPPSWIKV 2084
QY 2093 DLLAPMLIHGKITQGARQKFSLSIYISQITIMVSLDGKQVTVGNSGTGTLVYVFGNVDS 2152
DB 2085 DLLAPMLIHGKITQGARQKFSLSIYISQITIMVSLDGKQVTVGNSGTGTLVYVFGNVDS 2144

```

QY 2153 GIKENFNPIIARYIRLHPTHYSIRSTRMELMGCDLNSCMPLGMSKAI SDAQITAS 2212
D 2145 GIKENFNPIIARYIRLHPTHYSIRSTRMELMGCDLNSCMPLGMSKAI SDAQITAS 2204
QY 2213 SYTNMFPATWSPSKARHLQGRSNAMPQVNNPKWLOVDFQTKMKTGVTGKSLLT 2272
D 2205 SYLSSMLATWSPSKARHLQGRSNAMPQVNNPKWLOVDFQTKMKTGVTGKSLLT 2264
QY 2273 SMYVKEPLISSQDGHQWTLFFQNGKVKVFGQKODSETPVNSLDPLLLRYLRHPOSW 2332
D 2265 SMYVKEPLISSQDGHQWTLFFQNGKVKVFGQKODSETPVNSLDPLLLRYLRHPOSW 2324
QY 2333 VHOIATMEVLGCEAQ 2348
D 2325 AHIAIRLEVLGCDTQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
ID O62730 AC O62730:
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Factor VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney, and Spleen;
RA Gordy P.W., Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 2 FS/8 TYPE C DOMAINS.
DR EMBL; AF049489; AAC05384.1; -.
DR HSSP; P00451; 1CFG.
DR GO; GO:0005507; P: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58_C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CRC64;

Query Match 76.0%; Score 9439; DB 6; Length 2343;
Best Local Similarity 76.7%; Pred. No. 0;
Matches 1808; Conservative 199; Mismatches 325; Indels 24; Gaps 12;

QY 1 MQBELSCFFLCFLRPFCSATRRYVLGAVELSWDYMOSD-LGELPVDARPPRPVKSPFP 59
D 1 MQVELYICCFCLLPFLSLSATRKYLGLAVELSWDYMOSDLSALHADTSFSSRVFGLPL 60
QY 60 NTSVYVKTLFVEFTDLFNFAKPRPPWMLGGFTIQAEVYDVTIVILKNMASHPVSLHA 119
D 61 TTSVYKTVKTVFTDDLFNIAKPRPPWMLGGFTIQAEVYDVTIVILKNMASHPVSLHA 120
QY 120 VGSYVWASGAEAYDDQTSQSEKEDDKVFPQGSHTYVQVLKNGPMASDPLCLTYSYLS 179
D 121 VGSYVWASGAEAYDDQTSQSEKEDDNVFPQGSHTYVQVLKNGPMASDPLCLTYSYFS 180
QY 180 HVDLVKDLNSGLIGALLVCKSGSLAKKKTOTLHKFILLFAVFDGKSWHSPTKNSLMODR 239
D 181 HVDLVKDLNSGLIGALLVCKSGSLAKKKTOTLQEQFVLLFAVFDGKSWHSPTKNSLMODR 238

```

```

QY 240 DAASARAWPMHFWNGVYVNRSLVPLIGCHRSYVWHVIGMGTTPVHVSIFLEGHTFLVRN 299
D 239 ----AAQAQHELHFWNGVYVNRSLVPLVCHRSYVWHVIGMGTTPVHVSIFLEGHTFLVRN 294
QY 300 HRQASLEISPIITLTATQTLMDLGQFLSCHISSHQHDGMEAYVVKVDSCEEPQLRMKN 359
D 295 HRQASLEISPIITLTATQTLMDLGQFLSCHISSHQHDGMEAYVVKVDSCEEPQLRMKN 354
QY 360 EAEEDYDDDLTDEMDVVRPDDNSPSFQIRSVAKKHKTWHTYIAAREDDYAPLVL 419
D 355 ED-KYDDGLYGDMDVWSPDDSSSFFQIRSVAKKHKTWHTYIAAREDDYAPSGP 413
QY 420 APDRSVKSOYLNNGPQIRGKYKVRFMAYTDETFKTREAIQESGILGPLLYGEGYDT 479
D 414 TPNDRSHKLYLNNGPQIRGKYKVRFMAYTDETFKTREAIQESGILGPLLYGEGYDT 473
QY 480 LLIIFKQASRPNTYPHGTTDVRPIYSRRLPGVKVXELKDFPILPGEIFYKKTWTVVVDG 539
D 474 LLIIFKQASRPNTYPHGTTDVRPIYSRRLPGVKVXELKDFPILPGEIFYKKTWTVVVDG 533
QY 540 PTKSDPRCLTRYVSSFWNMBERDLASGLIGPLLI CYKESVDORGNOIMSDKENVILFSVD 599
D 534 PTKSDPRCLTRYVSSFWNMBERDLASGLIGPLLI CYKESVDORGNOIMSDKENVILFSVD 593
QY 600 ENRSWYLTENIQRFPLNPAGVQLEDPEFQASNMHSINGVYVFDLSQISVCLHEVAYWYL 659
D 594 ENRSWYLTENIQRFPLNPAGVQLEDPEFQASNMHSINGVYVFDLSQISVCLHEVAYWYL 653
QY 660 SIGAQTDFLSVFPFGVTFKHQVYEDTLTLPFSGGETVFMVSMENPGLWILGCHNSDPNR 719
D 654 SVGAQTDFLSVFPFGVTFKHQVYEDTLTLPFSGGETVFMVSMENPGLWILGCHNSDPNR 713
QY 720 GMTALLKVSSCDKNTGVDYEDSYEDISAYLLSKNNAIPEPSFQNSPHSTROKOFWAT 779
D 714 GMTALLKVSSCDKNTGVDYEDSYEDISAYLLSKNNAIPEPSFQNSPHSTROKOFWAT 773
QY 780 IPENDIEKTDPPFAHRTMPKIQNVASSDILLMLLRQSPHPGLSLDQAKYETTFSDDP 839
D 774 TPENDIEKTDQSGERTQLIKASVSSDILLMLLRQSPHPGLSLDQAKYETTFSDDP 831
QY 840 SPGAISNNSLSEMTFRPOLHSGDMVFTPEBSGLQIRLNEKLGTTWATBLKLDKPKVSS 899
D 832 SRGAIERNKGPPEVASLRPELRHSEDRFTPEPELQRLNENLGTNTVTLKLDLKISS 891
QY 900 TSNNLIS--TIPSDNLAAGTNTSSIGPPSPMPVYDSQDQTLFGKSSPLTSSGGPLSL 957
D 892 SSLSLWISPTIPSDKLAATEKTGSLGPPNWSVHFNGHLGTIVFGNNSHLIGSGVPLEL 951
QY 958 SEENDSKLLESLGIMNSQESSWGKNVSSYTESGRLFKGRAGHPALLTKDNALFKVSISSL 1017
D 952 SEENDSKLLEAFIMNIQESSLRENVLSMESNLFKEERIRGFASLIKDNALFKVSISSV 1011
QY 1018 KTKNTSNNSATNRKTHIDGSLIENSIPSWQNI--LESDTFKKVPFLIHDRMLMDKNAT 1076
D 1012 KTRNAPVNLTRNKRVAIPFLLIENSIPSWQIMLERNTFEFKVTSIHNFTFMDRNT 1071
QY 1077 ALRLNMSNKTTSKNMNVQCKKEGIPDDAQNPDMSPFKMLFLPESABWIORTHGKNS 1136
D 1072 ALGLNHSNKTTSKNMNVQCKKEGIPDDAQNPDMSPFKMLFLPESABWIORTHGKNS 1127
QY 1137 LNSGQGPSPQLVSLGPEKSVQGNFLSEKNKVVVGVGGEFTKDVGLKEMVFPSSRLFLT 1196
D 1128 LSSEQRPSQPLTSLGSEKSVKQNFLESE--KVVVGEDFTKDTQLQE--IFPNKSKIFFA 1185
QY 1197 NLNHLNHNTHNCKEKLQBEIEKETLIQENVVLPQHTVTGTXNFMKNIPLISTQNV 1256
D 1186 NLNHLNHNTHNCKEKLQBEIEKETLIQENVVLPQHTVTGTXNFMKNIPLISTQNV 1245
QY 1257 GSYDGVAVYLPQFRSLNDSTNRKTHAIFSKVGBEENLEGLGNQTKQIVKVACTRI 1316
D 1246 GLEBQYTPILQDTRSLNDSPHSEGHMANFSGIRREANLEGLGNQTNQWVFPSTTRM 1305

```

```
QY 1317 SPNTSQNFVTRQSRKALQFRRLPLETELEKRIIVDDTSTQMSKMKHILTPSTLTQIDY 1376
Db 1306 SSNAS-CHVITQGRKSLQKPRLSQSGHIKPERKVIANDTSTQMSKNNYLAQGLTLQIEY 1364
QY 1377 NEKEKGAITOSPDLSDCLTRSHSIFQANRSPPIAKVSPFSIRPIYITRVLPQDNSSHLP 1436
Db 1365 NEKEKRALITOSPDLSDCLTRSHSIFQANRSPPIAKVSPFSIRPIYITRVLPQDNSSHLP 1424
QY 1437 AA-----SYRKXDSGVQSSSHFLQAKKNNLSLAILTEMTGQDQREVSGSLGTSATNSVYIK 1492
Db 1425 ASACNYTFRETSVQVSGSHFLQAKKNNLSLAFVLTIGTEGQKFSLSGKSATNQPMYK 1494
QY 1493 KVENTVLPKDPDLPTKSKVLLPKVHYIQKDLPTFTSTNSGSPGHLDIVGSLLOGTEGAI 1552
Db 1485 KLENTVLPQGLSTSDKVELLSQVHDQSDSFPTKTSNDSPGHLDMGXIFLQKTCQPV 1544
QY 1553 KWNENANPCKVPPURVATESAKTPSKLPLPLANDNHVGTQIPKEEWSKOEKSPKTAEX 1612
Db 1545 KANKTNSPGKVPFKWATESSEKIPSKLGLVANDNHVDTQIPSEEWKSKOKSQNTATFX 1604
QY 1613 KKTILSLNACESHAJAAINEGQMKPEIYVIAKQGRTERLCSQNPVPLKRRHQREITRT 1672
Db 1605 RKDTILPLGPCENNDSAAINEGQDKPQRZAMWAKQCEPGRLCSQNPVSKHHQREITVT 1664
QY 1673 TLQSDQREIDYDDTISYEMKEDPDIDYDENQSPRSFQKTRHYFTAAVERLWDYQMS 1732
Db 1665 TLQDEEDKFYDDTISYEMKEDPDIDYDENQSPRSFQKTRHYFTAAVERLWDYQMSR 1724
QY 1733 SPVLRNRAQSGSVQPKVQVQFQFTDQSGFTQPLYRGELNEHLGLLGPYIRAEVEDNIV 1792
Db 1725 SPHLNRNRAQSGSVQPKVQVQFQFTDQSGFTQPLYRGELNEHLGLLGPYIRAEVEDNIV 1784
QY 1793 TFRNQAASRYSFSSLSYBDDQQAEPKRNFKVKNETKTYFMKQVHMAPTKDEFDCK 1852
Db 1785 TFRNQAASRYSFSSLSYBDDQQAEPKRNFKVKNETKTYFMKQVHMAPTKDEFDCK 1844
QY 1853 AWAYFSVDLEKDVHSLGLLGLLCHVNTLNPAHGRQVTVQEFALPTIIDEKSKWTFTE 1912
Db 1845 AWAYFSVDLEKDVHSLGLLGLLCHVNTLNPAHGRQVTVQEFALPTIIDEKSKWTFTE 1904
QY 1913 NMBRNCRAPCNIQMEDPTFKENYFHAINGYIMDTPLGLVNAQDORLWYLLSGMSNEI 1972
Db 1905 NMBRNCRAPCNIQMEDPTFKENYFHAINGYIMDTPLGLVNAQDORLWYLLSGMSNEI 1964
QY 1973 HSHFSGHVTVRKKEYSKVALYNLYPGVETVEMLPKSAKGIWKECLIGELHAGMSTL 2032
Db 1965 HSHFSGHVTVRKKEYSKVALYNLYPGVETVEMLPKSAKGIWKECLIGELHAGMSTL 2024
QY 2033 FLVYSNKCQPLGMAASHIRDFQITAGSQXGOWAPKLARLHYSGSINAWSTKBPFSWKV 2092
Db 2025 FLVYSNKCQPLGMAASHIRDFQITAGSQXGOWAPKLARLHYSGSINAWSTKBPFSWKV 2084
QY 2093 DLLAPMIHGIKQARQKFSYISQFIIMYSLDGKKWQTYRGNSTGTGLVFPFNVDS 2152
Db 2085 DLLAPMIHGIKQARQKFSYISQFIIMYSLDGKKWQTYRGNSTGTGLVFPFNVDS 2144
QY 2153 GIKHNIENPILARYILHPTHYSIRSLRMLMGLDCLNSCPMLGSKAISDAQITAS 2212
Db 2145 GIKHNIENPILARYILHPTHYSIRSLRMLMGLDCLNSCPMLGSKAISDAQITAS 2204
QY 2213 SYFTNMEATWSPKARLHLQGRSNAPQVNNPKWILQVDFQKTMKVTGVTQGVKSLLT 2272
Db 2205 SYFTNMEATWSPKARLHLQGRSNAPQVNNPKWILQVDFQKTMKVTGVTQGVKSLLT 2264
QY 2273 SMYVKEFLISSQDGHQWTLFPQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSW 2332
Db 2265 SMYVKEFLISSQDGHQWTLFPQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSW 2324
QY 2333 VQIOTLRMEVLGCEAQ 2348
Db 2325 AHHIALRLVLEIGCDTQ 2340
```

```
RESULT 3
Q7TN96 PRELIMINARY; PET; 2258 AA.
AC Q7TN96;
DT 01-OCT-2003 (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Factor VIII.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mistar; TISSUE=Liver;
RA Watzka M., Geisen C., Seifried B., Oldenburg J.;
RT "Sequence of the rat factor VIII cDNA.";
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY362193; AAQ21580.1; -.
SQ SEQUENCE 2258 AA; 251299 MW; 13AF91C788059B1D CRC64;

Query Match 50.6%; Score 6293.5; DB 11; Length 2258;
Best local Similarity 53.2%; Pred. No. 0;
Matches 1272; Conservative 317; Mismatches 629; Indels 171; Gaps 29;

QY 1 MQIELSTCFCLLRFCFSATRYVYLGAVELSDYMQSDL-GELPVDARPPPPVPSFPF 59
Db 1 MRAARGLCFLSLCALCSNATRYVYLGAVELPVDYVSGASGAGQSRSPPTPAP-- 57
QY 60 NTSVYVKKTLFVETDHLFNIAKPRPPWMLGLPTQAEVYDTPVITLKMASHPVSLHA 119
Db 58 SAHVHTKTVFVYMDRFHTAKPRPLMWGLLPTITVEHDTVITLKMASHPVSLHA 117
QY 120 VGSYVWKAASGAYDDOTSQREKEDKVPGGSHYVQVLKENGPMASDPLCLTYSL 179
Db 118 VGSFWKASGAYDDHSSPAEKDDKVLPGESHTYAMQVLGSGPMASDPLCLTYSL 177
QY 180 HYDLVKDLNGLIGALLVCRGSLAKEKTLHKFILLFAVDEGSKSWHSETKNSLMQDR 239
Db 178 HYDLVDRDLNSGLVALLVCKEGELSAERTMPPEFVILLFAVDEGSRSHAAT--R 230
QY 240 DAASAKAWPKMTVNGVYVRSPLGLIGCHRKSYVHVHGMGTTPEVHSIFLEGHTLV 299
Db 231 DPASTEAAQTVAHVNGYVNTLGLTGCRTSYVHVHMAVGTTPDIESILLEGHTSV 290
QY 300 HRAQSLIEIPITELTAQTLIMDLGQFLLSCHSSHQDGMAYVYKVDSPPEPOLMK 359
Db 291 HRAQSLIEIPMTLLTAQTLMDLGRFLLSCHSSHQDGMAYVYKVDSPPEPOLMK 350
QY 360 EAEYDDDLTDSEMDVVRPDDNSPFIQIRSVAKKPKTWVHYIAABEEDWDYAPLV 419
Db 351 EDTEDYDD--LDSEMDVFTWDPDAAP-FVQVRSAAKRPKTPVHYIAABEEDWDY 407
QY 420 APDRSVKSOYLNNGPQRIGRYKVKYFMAAYTDETKTEAIOHESGILGLLYGV 479
Db 408 TLEDGSRNRYGRGPRIGSKYKVFAYTDTGTEHTREATPREAGLGLLYGVGDS 467
QY 480 LLIIFNQASRPNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTV 539
Db 468 LLIVFNRRASRAVNIYPHGIRDVGAVHAGELPGEVGVKDLPIRPGSTKYRW 527
QY 540 PTKSDPRCLTRYYSVNNMERDLASGLIPLICYKESVDQRGNQIMSKRNVLFSVD 599
Db 528 PARSDARCVTRYASAVDPERDLASGLIPLICYKESVDQRGNQIMSKRNVLFSVD 587
QY 600 ENHSWYLTENICGFLENPAGVLEDEPFOASINMESINGVYFDSLSVCLHEVAW 659
Db 588 ENHSWYISNNRRPLDRAHVQLQDPEFASINMESINGVYFDSLSVCLHEVAW 647
QY 660 SIGAQTDFLSVFPSSGYTFXHKMVEYEDTLTLFPFSGTFVMSMENPGLWILG 719
Db 648 SVGAQTDFLSVFPSSGYTFXHKMVEYEDTLTLFPFSGTFVMSMENPGLWILG 707
```



```

QY 720 GMTALLKVSSCDKNTGSDYEDSIDISAYLLSKNNAIEPSSFSQNSRHPSTROKQFNATT 779
Db 708 GMTALLKVSSCDGVSYSYGETYGVLA-PiANDNAVDPSPSQNSHLHSRKKSN- 763
QY 780 IPENDIEKTDPAHRTTPNPKI QNVSSDDLMLLRQSPTHGLSLDLOBAKYETSDDP 839
Db 764 -----TPM-----RFLSHPTQGSFWSDSQGDVHDVHRAH 795
QY 840 SPGAIDNNLSSEMTFRPOLHSGDMVFTPESSLOLRNLKXLTAAATLKLDPKVS 899
Db 796 SPAAASHNEGPAQAQLPCHGPKKALPSS---WLAARSLVTTTWEAKKXLDLQVQ- 851
QY 900 TSNLLSTIPSDNLAAGTNTSLGPPSMFV-----HYDSQLDTTLTFGKKSSPLTE 950
Db 852 -----VSGLPD-----DRTTAVAPDDPVAACKKAGSSGFPDESSPAALGKKMYPRIR 899
QY 951 SGGPLSISENNDSKLLSEGLMNSQBSGKNSVSTESGLRFLPKGRAGHALLTKONALF 1010
Db 900 FHGFLSIIEGNRDSNSDSTLMYRLGSPPGDATSWTENRGLGKGRSHRVAFLARGNTLL 959
QY 1011 KVISILLKTKTNNKATNEKTHIDGPSLLIENSPPVWQNILESDETFKXVTLPHDRML 1070
Db 960 -----SDVDEKSHAPGPTSVGNSTAAVQDTILICSBILEVTPIIDRIL 1004
QY 1071 MDKNAALRLNMSKNTTSSKNMVMQOKKEGPIPPDAQNPDMSPFKMLFELPESARWIOR 1130
Db 1005 SDAKATYLRPHRTPDRTITSTERKDIRHEKDGDLVPQDADDTSAFPSEAPLSSTDWLKE 1064
QY 1131 THGKNSLNSQGPSQPLVSLGPEKSVEGQNFISEKKNVVGVGEBTKVDGLKEMVPPSS 1190
Db 1065 ANGDNKAKBPQESPQLVLYLMYIKMENQSPSEKKNKVIAGQGGETKNTGLEDTVPFRK 1124
QY 1191 RNLFNLNLMHNHNENQEKKEKKEKTELIOENVLPLQHTVTGTKNVKNLFLLS 1250
Db 1125 TSVFLTTVAKQBSGRHQB-NIPQAVTKEAIEIEKALPQVHIAATGSKNFRDMFVLG 1183
QY 1251 TRONVGSVDGAVPVLQDPRNSLNDSTNRKTKHTAHFSKKEGEE--NLBGLGNQTXIVE 1308
Db 1184 TQONI-NLHEETVVPVHKVWPTRNPTDTRQIPMVHFFKRRKEETNNGGLVANKTRETVR 1242
QY 1309 KYACTTRISPTSQQNFVQSRKALKQPLPLEETELKRIIVDTSTQKSNMKHLP 1368
Db 1243 NY-----PSQKNSVARRQOAGRIK-----ASARWLPDVNRSIQ 1277
QY 1369 STLTQIDYNEKEKGAITQSPSLCLTRSHSIPQANRSPPIAKVSSFPSPRIPLVLRVL 1428
Db 1278 SLLKQIDHRKERKFIIESRADS-SGTKSPQTNSHSPHVWVMSAPP---PADIRIRS 1333
QY 1429 QDNSSHLPAASY-----RKXDSGVQBSHFLQGAKKNLGLAILTLEMTGDQREVGSLGTS 1484
Db 1334 RD-SSQVWVSSVAYDPTGSSRIESSPFLNETETDNPSLAVPPRFRVRRGRFASPEKV 1392
QY 1485 ATMSVYKVKVENTVLPKPLPKTSKGVKELLPKVHIYQKDLFPNETNGSPGHLDLVEGSL 1544
Db 1393 NTHSACDPENNVSEPVSPGBAVTVALPPHVSTQEEPLPAGSSCERRGHVDLFLETS 1452
QY 1545 LQTEGALKWNEANRPGKVFLEVATESSAKTPSKLLDPLANDNHYGTQIPKEWKSQEK 1604
Db 1453 LQTRGFEVRSRRSGG-----DTEGHTENPGTKTSPSL-----PQMPKQDQSGEAG 1499
QY 1605 SP-----EKTAPKKDITLSLNACSNHAIAINQGNQKPEIETVWAKQGRTERLCSON 1658
Db 1500 YPKICSRWGDGTVLPFRPHNLSLGA-----KKKNPLPREATRVQEGEAPSLVSPE 1549
QY 1659 PPVLKXHQREITR-----TTLQSDQREI--DYDDTISVENKXEDPDIDYB 1701
Db 1550 PLVLRPRPREASTLLPGGEMREDDGCVTAEDYDDDDVMAEYDDAVVTDTEDFDIYGE 1609
QY 1702 DENQSPRSQKTRHYFIAAVERLMDYGMSSSPHVLNRNAQSGVPOFKKXVFOERTDGS 1761
Db 1610 EAGQGRGFGQKTRHYFIAAVERLMDYGVVRSFGALGDRANSGDAARFKVFRFRTDGS 1669
QY 1762 FTQPLYRGELNHLGLLGYIRAEVEDNIVMTFRNQASRPYSFYSSLISYEEDRQGRAP 1821

```

```

Db 1670 FTQVRHGERGLDHLGLLGYIRAEVEDNIVMTFRNQASRPYSFYSSLVSPED- GGAAP 1728
QY 1822 RKNFVKNPTKTYFWKVOHHMAPTKDBFCKAWAYFSDVDLEKDVHSLGLLPLLVCHTNT 1881
Db 1729 RSNFVRNPTKTYFWRVRFHMAPTDGEFCKAWAYFSDVDLEKDVHSLGLLPLLVCHAST 1788
QY 1882 LNFAGHROVTVQEFALFFETIPBETKSWYFTENMERNCRAPCNTQMBDPPFKENYRPHAIN 1941
Db 1789 LHESHARPLAVQBFALLFAVFDETKSWYCAENLDKRCRPRGTPAGDPARWEYRPHAVN 1848
QY 1942 GYIMDTLPLGLWAQDORIRWYLLSMGNSNENIHHSFSGHVFTVRKKEEYKVALYNVYPGV 2001
Db 1849 GYVADALLGLVWAGERTRWHLLSMGDPGHAQSVHPSAHSVTVRDGGEHTAVCNVYPGV 1908
QY 2002 FETVEMLPKAGIRWRVECLIGEHLHAGMSTLFLVYSNKQOTPLGMASGHIRDFOITASGQ 2061
Db 1909 FTTVEMLPKAGIRWRVESLVGEHLRAGMSALFLVYSTRQVPLGMASGYIRDSQITASGY 1968
QY 2062 YGOWAPKLARLHVSYSINAWSTKEPESWIKVDLLAPMILHGIKTQAGQKQFSSLYISQFI 2121
Db 1969 HGWTPLQARLHVNAGSVNWSAKEPPFAWIKVDLLAPMILHGIETQGARHRLSSLYVSQFI 2028
QY 2122 IMYSLDGKKWQTYRGKNTGTLMVFFGNVSSGIGKHNIFNPPIIARYIRLHPTHYVIRSTL 2181
Db 2029 IMYSLDGKQWLSYRGKNTGSLVFFGNVDASTVVRHNFPPPIVARYIRVHPHTASIRTAL 2088
QY 2182 RMELMCCDLNSCMPUGMESKAISSAQITASSYFTNMFTWSPSKARLHLOGRSNAWRPQ 2241
Db 2089 RMELMCCDLNSCAPLGMSGWIISDTQVTASSHLSTFSANPPALARLELRGGANAWRPQ 2148
QY 2242 VNNKEMLOWDKTKMKVTVGTQGVKSLTSMYVKEPILSSQDGHQWTLFPQNGKVKV 2301
Db 2149 VADPTQWLQVDQRTVVKVTVGTQVARSLLTAMFVKKFLVSTSDGRHHTHVLQDGKVKV 2208
QY 2302 FQGNQDSFTPVVNSLDPPLTRYLRHPQSWHVOIALRMEVLGCEAQDL 2350
Db 2209 FQGNRDASTPMVNSLHPFRTRYLRHPQVWEQIALRLEILGCEAQDL 2257
RESULT 4
Q804W6 PRELIMINARY; PRT; 1639 AA.
AC C804W6;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Coagulation factor VIII precursor.
GN F8.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Ial K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF465279; AAC33374.1; --
DR GO; GO:0005507; F:coppper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 2.
DR Pfam; PF00754; FS_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.

```

not
prior Art

D5	1296	LVETVGEQSLGMRKALVYNPQCSERPLGKSGRIGDSQIKASDVIGNWLPRLARLDS	1355
QY	2076	GSINAWSTKEPESWIKVDLAPMIHGIKTQARQKFSYISQFTIMTSLOKKWQTYR	2135
D5	1356	GYINAWMGYNEKSWIQVDLCAPTLHRVQTQGVRSNLRNNTYIATFVSYSLPQETWSTYR	1415
QY	2136	G-----NQTGLTMVFFGNVDSGIGNLFPNPIIARYIRLHPHTHYSIRSLRMLMGCD	2189
D5	1416	GSSSSSSSSSSTAKVGNLDSNRVKNPFPVFPVARYIRIHPLYNQRPALRMELMGCD	1475
QY	2190	LNSCSMPLGMSKASDAQTASSYTNMFWATWSPSKARLHLCGRNANRPQVNNPEWL	2249
D5	1476	LNSCSPLGLQDRRIPEDSFVASSYSWLLRSWTPSLARLHOGSANAARPKNNPHEWL	1535
QY	2250	QVDFQKTYKVTGTYTGGVKSILTMVKSFLSSODGHQWTLFPQNG--KYKVFQGNQD	2307
D5	1536	QVDLGKVKKAITGVVTCGARSLLTKMVTBFSVTISRDQAWSSVLGSSQREXIFQGNND	1595
QY	2308	SFTFVYVNSLDPPLLTLYRLHPQSWHQIALRMEVLGCEAQ	2348
D5	1596	SDBEALTIFDAPLFGYRIHPHGLWINDIALRLEVLGCDTQ	1636
RESULT 5			
Q90X47	ID	Q90X47	PRELIMINARY; PRT; 2119 AA.
AC	Q90X47		
DT	01-DEC-2001	(TEMBLrel. 19, Created)	
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)	
DT	01-OCT-2003	(TEMBLrel. 25, Last annotation update)	
DE	SC:B22015.3	(Novel protein similar to vertebrate coagulation factor V and VII).	
DE	SC:B22015.3		
GN	Brachydanio rerio (Zebrafish) (Danio rerio).		
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
OC	Cyprinidae; Danio.		
OC	NCBI_TaxID=7955;		
RN	[1]	SEQUENCE FROM N.A.	
RP	Lloyd D.;		
RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
RC	-!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
DR	EMBL; AL590146; CAC94896.1; "		
DR	GO; GO:000507; F:coppper ion binding; IEA.		
DR	GO; GO:0007155; P:cell adhesion; IEA.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR008972; Cupredoxin.		
DR	InterPro; IPR000421; FA58 C.		
DR	InterPro; IPR008979; Gal Bind like.		
DR	Pfam; PF00394; Cu-oxidase; 2_		
DR	Pfam; PF00754; F5_F8 type C; 2.		
DR	SMART; SM00231; FA58C; 2.		
DR	PROSITE; PS01285; FA58C_1; 2.		
DR	PROSITE; PS01286; FA58C_2; 1.		
DR	PROSITE; PS50022; FA58C_3; 2.		
DR	PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.		
QY	SSQUENCE	2119 AA; 240643 MW; DCOE806FFA8761E6 CRC64;	
Query Match			
Best Local Similarity 30.3%; Pred. No. 2e-159;			
Matches 752; Conservative 379; Mismatches 825; Indels 529; Gaps 68;			
QY	11	LCILRR-CFSAT--RYYILGAVELSDVYMQSDLGELPVDARPPRPVKSPFPNTSVYK	66
D5	13	LALLAFILCHCATVERHYIAANINMDYTSQQO-----RTGQSYK	53
QY	67	KTLFVETDHLFIARPPPMWGLGPTIOAEVYDTVITLKNMASHPVSLHAGVSYWK	126
D5	54	KVYREYNEG-FKQKAPHLSSGLLPTLRGQCDTIIVTFRNWADHPCSLHPHGAYGK	112
QY	127	ASGAEYDDQTSOREKDDKVFPGGSHYVQVLEKNGPMASDPLCLTYSYLSHVDLVKD	156

D5	113	QSEGSYFDNTSLLKNDVDVIOPEGEHTYQWDVTSDVITTAADPPCITTYLSHPDIVD	172
QY	187	INSLIGALLVOREGSLAKBKQTTLH---KFILLFAPVDEGKSWHSETKNSLMQDRDAAS	243
D5	173	YNTGLIGPMILCKGFTLDDSGNQ-IHFQESVLLFLGVFDENKSWYS-----TGDS	221
QY	244	ARAPMCHTIVNGYVANSPLGLIGCHRKSVVHVIGMTTPEVESIFLEGHTFLVRNHQA	303
D5	222	POPLNVKITYINVTNGSVFDDLDICAHKSVSWHLLGMSSEPELFSVHFNQVILLHCHKTS	281
QY	304	S-EISPIITFLTAQTLIMDLGQFLASCHTSSHOHDGMEAYVKVDSCEP--EPOLRMKNNEE	361
D5	282	AVGIISGTATASMTGVHPGRWLNGSHISKLEAGLGLYLNIRKDCDYATAPKRLLTIEQ-	340
QY	362	ABEDYDDDLTSDMDVVRFDNNSPSFIQIRSVAKKHPTWVHYIAAREEDWDYAPLVLAP	421
D5	341	-----KKESQEWTYMAAEVIMDYAPNMPEN	367
QY	422	DDRSYKSOYLNNGPQIRGKYKVRFMAYTDETFKTR---EAIQHSGLIGLPLYGEVGD	478
D5	368	MDGDFRSKYLKQGPQIRGKYKAVFTQYKQGMFKERABDKORKRELGIIGPVIRAIQID	427
QY	479	TLIIIFKNOASRPYNIYPHGIITDVRPLYSRRLPKGVKHLKDPFILLPGEIFKYKWTIVED	538
D5	428	IIKIVFKNKASRPYSIYPHGLTIIDKAAEGASYPQGN--QTVSVQPGETYTYTWSVTEED	485
QY	539	GPTKSDPRCLTRYSSFYVNMERDGLASGLICPLLICYKESVDQDQGMQIMDKENVILFSVP	598
D5	486	VPTSDPRCLTMTMSAVDAPRDIASGLVGPLLICKSOLNKKVQLKADKQOHAMFTVF	545
QY	599	DNRSWYLTENIQRLPNPAGVQLEDPEPQASNMHSINGVYVDSIQ-LSVCLHEVAYWY	657
D5	546	DENKSYQDENINTVCSDPKVKKDDPEFYKNVWHTINGVYVESQELGFCHEIVTWH	605
QY	658	ILSIGAQDTFLVFFSGYTFHKMYVEDTLTFPESGETVFMMSMENPGLWILGCHNSDPR	717
D5	606	VSSVGEQDYQIQTATFYGHTFELKNEEDILSLFPWTGETITKMMVNIIGIWLASLNHSDS	665
QY	718	NQMTALLKVVSSCDKNTGDY-YEDSVEDISAVLLSKNAIEPRSFQNSRHSSTROKQFN	776
D5	666	TKGMVKFKDLECFR---DYIEYDIED-----GKFTAWKPTI-----	701
QY	777	ATTIPENDIEKTDPMFAHPTPMFKIQNVSSDILLMLLROSPTPHGLSLSLDQAKYETFS	836
D5	702	-----NEIKKEEPEVAR-----	713
QY	837	DDPSGAIDSNLSLSENTHFRPQLHHSQDMVFTPESGLOLRINEKIGTTAATLKLKLDPK	896
D5	714	---PDVDEYSDLFAET-----LNLRTFNKVKDEVEIDLTL	746
QY	897	VSTSNLLISTIPSDNLAAGTNTSSLGPPSPVHVHVSQDITTLFGKKSPLTESGGPLS	956
D5	747	FLDQDDGLLPIVEEKSLGSSNEN-----LHNATLQSFIEHNG-----	783
QY	957	LSENNDSKLLSEGLMNSQESSGKNVSVSTESGRLPFKGRAHGFALLTKDNALFKVISL	1016
D5	784	-----LLMEEGDLDKGESSKNVLNDSTDK-----AL	809
QY	1017	LKTNKTSNNS---ATNRKTH---IDGPSL--LIENSFSPVWQNLLESDDTEFKKVTPLIHOR	1068
D5	810	LETTTTFDSNRVVALNNETDSILDFPIVERKVSAPSKPMNEPESTVWTFKTEHI---	866
QY	1069	MLMDCKATALLRHNHNSKTTSSKNMVMQKKGPIPP-DAQNPDMSF-----FKMLFLP	1122
D5	867	-----NSSLERINAIYSPITETNINTMTHTDTPSTIPDGGSTGEMNFTLEDDTALLSS	921
QY	1123	ESARWIQRTHGKNSLNSGGQSPFKQLVSLGPKXSEVGQNF-----LSEKNKYVVGK	1173
D5	922	ESEPLQSNQSENRIAFQBELNAKOGTDVDNNNSVKNQIFKYNVPBGDTLSNKKIQV--	979
QY	1174	GETFKDVGLEKMWFPSSRNLFITNLNDLHNHTHNOEKIQOEIEKKEKLIQSNVLPQI	1233

Db 980 -----BEDFVLLSSVFSEMS-----TMEYDVSQDITVKESKETAQS 1018
Qy 1234 HTVCTKFMKNLFLSTRQWVEGSDYCAVAPVLQD-----FRSLK-----DS 1276
Db 1019 QELSSTKTYSGEIIILSLPDIITAFNLSSSVLRNNSLESNESSNETLFWSSNATFSDS 1078
Qy 1277 TNRKKHT-----AHFKKG--EENLGLGNQTIQIVKXACTTRISPTSQQNFVQRS 1330
Db 1079 TNATSSPSSTATFADFNTTFSTNATFSDFSNRISQMSDSSNATLSDSSNAT-----LSDSS 1134
Qy 1331 KRALKQKRLP--EETELEKRIIIVDTSTQSKNMHGLTPS-----TLTQIDYN 1377
Db 1135 NATLSD-----SSNATLSSNATLSSNATLSSNATLSSNATLSSNATLSSNATLSS 1184
Qy 1378 -----EKEGGAITQSPLSCLTRSHSIPQANRSPILAKVSPFSPRIPIYLRVL 1427
Db 1-85 SNTTLDLSLESEMTYILSANDTI-KSHSEVVSNTSQL-----SSSESTENISL--L 1235
Qy 1428 FODNSSHLPAASYRKKSQGVSESHFLQCAKKNLSAILTLEMTGQREVGSIGTSATN 1487
Db 1236 YGSLN-----ASSMKNDSESESEEVYILNKNH-SEAILTSHL--DQKE-EHWGYESKH 1286
Qy 1488 SVTVKVENTV-----LKPDLKPTSGKV--ELLPKVHI-----YCKDLFTET 1529
Db 1287 ELVHKELPDHNNKYVKKSAANSNKPKEKKKVVQVFKPKYGMKTKSKDYKQPR 1346
Qy 1530 SNGSPGHLDLVGSLLQCTGCAIKWNEANRPGKVPFLRVATES-AKTPSKLLDLAWDN 1588
Db 1347 SPSFP-----RGFGSVLTGRSRP-----VSEDELTEKPIVIGVPRDFN 1388
Qy 1589 HYGTQIKPEEKSEKPEKTAFFKKKDTILSLNACENHAIANEGNQKPEIBVTWAKQ 1648
Db 1389 DYELIYIPKQDEAD-----FDGL-----DHPE----- 1411
Qy 1649 GRTERLCSQPPVLKRHQREITRTLOSQDEIDYDTISVEMKEDFDIYD-EDENQSP 1707
Db 1412 -----EYEVYKDPYS---KTADVQALDATSQHLLK 1440
Qy 1708 RSFOKTRHYFIAAVERLWYMGSSPHVLRNRAQSGSVPOFKKVPQEDTGSFTQFLY 1767
Db 1441 MAGDKNTRTYFISVEEBEDVYA-GYGORRLDKTAQNERPTVFRKVFERYLDSFIRDI 1499
Qy 1768 RGLNEHLGLIGPIRAREVDNIWTFNQASRYSYFVSSILSY-----EDQKQGA 1819
Db 1500 RGEDELHGLIGLPIKAEDVTWVFFRNARSYSUHGAVKYLKQWGLSYDDESFPY 1559
Qy 1820 BPRXKFKVNETKTYFWKVQHMAPTKDEDFCKAWAYFSDVDLAKDVHSGLIGLVLVCHT 1879
Db 1560 YKQDDAVPPNGTFYMTWINKPGSPQNESDCTWTYYSAWPERDINSGLIGLVLVCRK 1619
Qy 1880 NTLN--PAHGRQVTVQEFALFTTFDETKSWYFTENMERNCRAPCNTOEDPTPKENYRF 1937
Db 1620 GTLKKPEDRR-----BFVLLFMTFDEKSWLYEENRQRIERKRRVVM-DPNFQDLKFP 1673
Qy 1938 HAINGYIMDTLPLGLMAQDQIRWYLLSMGSMENIHSHFESGHVFTVRKKEKXVLYNL 1987
Db 1674 DALNG-IYSLKGRXYTNQAKHLLNMGSPKLLHSHVHGQTFINKELKDRHQGYIPL 1732
Qy 1998 YPGVFTEMLPSXAGIWRVECLIGEHLHAGMSTLF-----LVYSNK 2039
Db 1733 LPGAFFATLEMLPSKXGLWQLESEVGLSQGRNQTLFLFDESYYKRYVCQVLRSLYIFAV 1792
Qy 2040 QOTPLGASGHIRPQFOTASQYQCAWPKARLHYSSINAWST-KEPPSWIKVDLLAPM 2098
Db 1793 CDHPLGLISGTVQDEQITASTRCQWPHLARLHNTGKYNASTSPBPQVQLQVDFQFP 1852
Qy 2099 IHHGKTQGAQKPFSSLYISOFILMYSLDGKKWOTYVGNSTG-----LMWTF----- 2146
Db 1853 VISKATQGAQFTHHFLVNTTYSYDCKKWIYYKGDSDAVKRNHKTNVFFNARQT 1912
Qy 2147 --GNVDSGKIKHNFPPIIARYIRLPHPTHYISIRSLRVELMGCGLNSCMPLGMSKAI 2204
Db 1913 PEGNABAVETKENIFFPPLIGRYVRLHPLSHSYNFPVRLVLEYGCELGCCVPLGMEGLI 1972

RESULT 6

O88783 PRELIMINARY; PRT; 2183 AA.
ID O88783
AC O88783; PRELIMINARY; PRT; 2183 AA.
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Murine coagulation factor V.
GN F5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98282202; Pubmed=9616155;
RA Yang T.L., Cui J., Rehmutulla A., Yang A., Moussalli M., Kaufman R.J.,
RA Ginsburg D.
RT "The structure and function of murine factor V and its inactivation by
RT protein C".
RL Blood 91:4593-4599(1998).
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; U52925; AAC99553.1; -.
DR HSP; P12764; T42764.
DR HSP; P12764; T42764.
DR MGD; MGI:88382; F5.
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR008117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal bind like.
DR Pfam; PF00394; Cu-oxidase; 3_C-like.
DR Pfam; PF00754; F5_P8 type_C; 2.
DR SMART; SMART:SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
DR PROSITE; PS50022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
SQ SEQUENCE 2183 AA; 247228 MW; BFC8AA723F60317 CRC64;

Query Match

22.3%; Score 2774.5; DB 11; Length 2183;
Best Local Similarity 29.9%; Pred. No. 2.4e-156;
Matches 750; Conservative 408; Mismatches 851; Indels 503; Gaps 73;

Qy 8 CFFLCIL--RFC-----FSATRYLGAVELSDYMQSDLGHLVYDARPPRPVPS 56
Db 7 CFFLLVLTGRWAGWSHQAAQLKQFYAAQGLNWHPE-----PTD 51
Qy 57 FPENTSVYKTLFVFTDHLFNIAKPRPFWMGLLGTPTQAEYVDYTWITLKNMASHVPS 116
Db 52 PLSNSIPSPFKIIVREY-EQYFKKPRSNLSGLLGTLYAEVGVKIVHFRNKAADPLS 110
Qy 117 LHAVGSYKASGAEYDDQTSOREKDDKVPFGSGHYVWQVLKNGPMASDPLCLTYS 176
Db 111 IHPQIKYKSFSGASYADHTFPAERKDAVAPGEETVYEWIVSDESGTPDDPCLTHI 170
Qy 177 YLSEVDLVKDLNSGLICALLVCEGSLAKEKTQL--HKFILLFAVEDEGSKSHSETKNS 234


```

QY 2242 VNNPKRMLOVPQKMKVGTGVTQGVKSLTSMYVKRELISSQDGHQWTLFFQNGKV-- 2299
D5 2075 ANNKNQMLQVLLKIKKVTATVTOCKSLSSGMSYKSYIQVSDQGVAKVYRQKSSMVD 2134
QY 2300 KVFQCNQDSFPPVNSLDPBLTLRLTRHPSQWQHQIALRNEVLGCEAQDLY 2351
D5 2135 KIFEGSNKCKHMKMFFNPPIIRIPRIPTWQNSIALRLFLGC---DIY 2183

RESULT 7
Q804W5 PRELIMINARY; PRT; 1802 AA.
AC Q804W5.
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Coagulation factor V precursor (fragment).
GN F5.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tudenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes."
EL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465280; AAC33375.1; -.
DR GO; GO:000507; P:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal-Bind-like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00222; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
FT NON TER
FT 1
SQ SEQUENCE 1802 AA; 205160 MW; E2EC3D1D9AA836FD CRC64;

Query Match 21.6%; Score 2688; DB 13; Length 1802;
Best Local Similarity 30.4%; Pred. No. 2.6e-151;
Matches 710; Conservative 322; Mismatches 719; Indels 586; Gaps 67;

QY 62 SVVYKTLFVEFTDHLFNIAKPRPPWMLGLGFTIQAEVYDTWITLKNMASHPVSLHVG 121
D5 2 SPTYKKVVFREY-DKDRQPKASRPPWLLGLGPTLRAEEGEIIVTFNLTAKPYSIHPHG 60
QY 122 VSYWKAEGEAYDDQTSQREKEDKVPFGSHYVWVQVLEKNGPMASDPLCLTYSLSHV 181
D5 61 VAYGQSEGANFYDNTSQEKEDDVWVFNSEHYWYBITSDVSPQQNDPTCLTYIISHK 120
QY 182 DLVQDLNSGLIGALLVCREGSLAKEKQ--TLHKFILLPAVDFEGKSWHSTKNSLNQDR 239
D5 121 DVVEEYNSGLIGALLVCKSGLDSESGQIGIYHSEVFLFGVFENE-----SKPKQN 172
QY 240 DAASARAWPKHYVYNNRSLPGLIGCHRKSVYVHWIGMTTPEVHSIFLEGHTFLVRN 299
D5 173 DPASVD--HIKYTINGTEGSLPDVSICTYAPVSLHLVHGSSDDEVSVHNGOVLOQNG 230
QY 300 HQASLSISPTFTATLMDLQQLLSCHISSHQHDGMEAYKVDSCP--EPPQLRMK 357
D5 231 HKMSSVGLIGSSTVSMVAVHTGRWLLSSQIMKHIQGMVGFVNMBCIGCFKAPQRTLT 290
QY 358 NNEAEADYDDDLTDEMDVVRFDNDDNSFPTQISVAKKPKTWIHAEEEDWDIAPL 417

```

```

D5 291 REQ-----RKNREWTYIAAEIIVANTAPN 316
QY 418 VLAPDDSYKSOYLNNGPQIRGRYKVKVRPMAYTDTFKT---REAIQHSGLIGPLIYG 474
D5 317 EPAHIDQYKLYLRQSSSTRIGGKKAIVTLTNESFTQVHKKQKQKDELGILGPVIRA 376
QY 475 EVGDTLLIIFKNOASRPNIYPHGITDVRPLYSRRLPKGVKHLKDFILGEIKPKYKTV 534
D5 377 QIRDVIVIKVKAIRPYSIYPHGLTIKSEEGVNPYPGSNH--SHSVQPGETHYTWKV 434
QY 535 TVEDGPTKSDPRCLTRYSSFVNMERDLASGLIGPLIICYKESVDQGNQIMSKRNVL 594
D5 435 VEDEPLDDDDARCLTRYLSAVIDPFDIASGLIGPLIICKESLNVNRVQLRAKQEHAM 494
QY 595 FSVFENRPNWLTENIOFPLNPAGVQLEDPEFOASNMHSINGYVFDLSQ-LSVCLHEV 653
D5 495 FAVEDENKSWYLDNI--RYRCEHSKVNKADPDFYKSNWMSINGYVFESSPVLGFCNGEV 553
QY 654 AYWYILSIGAQDTFLSVFFGYTPKHKMWYEDTTLTLPFSGETVFMGMENPGLWILCHN 713
D5 554 ATWHVSSVGAQDTIQATFYGHTFELNERTEDFLSLYPMTGETISMMMDNIGWLLASLN 613
QY 714 SDFNRGWTALLKVSSCDKNTGYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROK 773
D5 614 SHGTTKGRVKFQDVECYD---YQVEYDD-----STNV 644
QY 774 QFNATTIPENDIEKTDWFPAHRTMPKIQNVSSDILLMLLRQSPFHGLSISLQOEAKYE 833
D5 645 EFNV-----W-----NPLSLDTIQE--- 659
QY 834 TFSDDPSFGAIDSNLSLSEMTFRPQLHSGDMVTFPESGLQLRLNEKLT-TAATLKK 892
D5 660 ---NEKPILVENKV--GEFDSYTEDLAN---ELGLRSLKQOSNSDVBQLDSF 705
QY 893 LDFKVSSTSN---NLISTIPSNLAAGTDNTSLGSPMPVHYDSOLDTTLFGKKSSPL 948
D5 706 LDYEVVDVSNETQINFTKT-KDNFNAABINGLN-----DT----- 739
QY 949 TSGGPLSLSENNKDSLLESGLMNSBESSGWKNVSTESGRLFKGRK-----HGP 1000
D5 740 -----SISNRLDVQNLTGADLLN-QQINWNTVMQNNSTLAPLAHNSVNEIKLMNHSP 791
QY 1001 ALLTKONALFKVSIILKTKNTSNATNKRKTHIDGFSLLIENSPPSVQWNLSDSTF-K 1059
D5 792 Q--TKNT--FSVIDTDLEAATVNTSA-----LSVDNYSK 824
QY 1060 KVTFELIHDRMLMDKNATALLRNMNSKTTSSK-NMENVQOKKEGPIPPDQNPDMGFPM 1118
D5 825 EITNLA--GTLOGKNLTA---SDSNVTSVSRINM-----PATRIP----- 859
QY 1119 LFLPESARWQRTQ--GKNSLNSGOGSPKOLVSLGPEKSVGQNFLEKKNVWVGKEFT 1177
D5 860 -----SCLSVSKAHVGN--NSSDGTNSSTLEAHGP-----YLNSSGV----- 896
QY 1178 KDVGLKEMVPPSRNLFNTLNLHNNHTHNOBKIQEEIEKKEKTLTIQENWLPQIHVT 1237
D5 897 -----NPTISKSGNVAALLKNGSVTVKLPMSKEEL----- 927
QY 1238 GTKNFMKNLELLSTRONVEGSDGAYAPVLDPRSLNDSTNRTKHTAHFKKGEENLE 1297
D5 928 -----DNTSLKNPSNKTSAESS----- 944
QY 1298 GLGNQTKQIVEXYACTTRISPTNSQNFVTRQSKRALKQFRUPLBETLEKRIIVDDTST 1357
D5 945 -----TPSVSTYNDIMTSSKE-----LSSSESSEKVEF----- 971
QY 1358 QWSKNMKHLTPSLTQIDYNEKEKGAITQSPLSDCLTRGHSIPOANRSLPLAKVSSFFS 1417
D5 972 -----IYVKDKKAGLI----- 982
QY 1418 IRFIYLTRVLFDQNSHLPAAASYRKQSGVQVSESH--FLQAKKKNLSAILTLENTGQ 1475

```

```
Db 983 -----KTSVKTSGHNNWYDGTGK-----IVSAEIPDDM 1011
Qy 1476 REVSGLSGTSATNSVYKKVNTVLPKPDLP-----KTSKVELLPKVHIYQKDLPTET 1529
Db 1012 KKFEM-SPQNKKTTRVN-----RHRQKGGMKTKRKYKPO-----PRSGPLSP 1061
Qy 1530 SNGSPGHLDLVEGSLLOCTGCAIKWNEANRP-GKVPFLRVATESSAKTSPKLLDPLAWN 1588
Db 1062 RGNP-----LMSPGA-----RPGQLQPVNN--DETLLNMPVVGVRPDPFS 1102
Qy 1589 HYGTOIPKEZKSOEKSPKTAFFKKDTILSLNACSNHAAINEQCPETEVTVWAKQ 1648
Db 1103 DYELPLGDE-----1112
Qy 1649 GRTERLCSQPPLKRRHOREITRTTLOSODEIDYDDTISVEMKKEFDIVDBENQSPR 1708
Db 1113 -----PDHLDVQRNVK-----ANEYEVNYKDPYSNEDAKNLHLHQTKYLENR 1158
Qy 1709 SFQKKTTHYFTAAVERLDWYGMSSPHVLRNRAQSGS---VPOFKKVVVFQETDGSFTQ 1765
Db 1159 --DKDVRTYFTAAVEQWDY-----AGYQRRRRRPPGQNRHTKPKVVRFLYLDSSRTP 1212
Qy 1766 LYRCELNEHLGLIGPYTRAEVUNIMVTPRNOASRPYSFYSLLSYEBDQOQA--BPRK 1823
Db 1213 EVRGEVDEHGLIGLPVKAEGVQITMVVFVKNASRPFSLHPNGVYSK-QTEGLSYEDGS 1271
Qy 1824 NF-----VKPNETKTYFWKVQHHMPTKDEPCKAWAYFSDVLEKDVHSGLIGLLV 1876
Db 1272 NYWKYDNEVQGNATFYIWNVPMVGPTDSENCRITWAYISGVNPERDIHSGLIGLLV 1331
Qy 1877 CHTNLNPAGHQVT--VOEPALFTTIFDETKSMYFTENMEENCRAPCNIQMEDPTPKENY 1935
Db 1332 CQKGTLN---QELTNTREEMLLFMTDESCSWYFDRNREIMLRNWX-KVMPDPIWENL 1386
Qy 1936 RFAINCYIMDTPLCLWAOORIRWLLSWGSENTHSHFSGHVTVTRK-BEYKMAL 1994
Db 1387 KFSINGIITN-LGURMYTNQLVSMHLNMGSPKDFNSVHFQGTTLHKKTKTSYQAV 1445
Qy 1995 YNLYPGVFETVEMLPKAGIRWRECLAGEHLHAGMTLFLVYGNKCTQPLGMASGHIRDF 2054
Db 1446 YPLLPGSFATLVMYPSKEGLWQLETEVGINCEKGYQTLFLVLADDCVHPGLRSGSVND 1505
Qy 2055 QITASGOYGQWAPKLARLHSGSINASTKBPFWIKVDLAPMIHIGITQGARQKPS 2114
Db 1506 QITAINTRGWEPHLARLHNGKTNASTQCNYSWIQWDFQRPVVISQVATQGAQKQFOA 1565
Qy 2115 LYISOPTIMYSLDGKKWQTYRGNSTGLMYVFFGNVDSSGKHNLENPITARYLRLPHT 2174
Db 1566 QYSSYVYISNDRSWSFYKGSRDDIKVFTGNNDPYDKKNTFFPPLIGRPIRHFPLK 1625
Qy 2175 YSIRSLRMLMGCDLNSCSPMGESKAISDAQITASSYFTNMFA-TWSPSKARLHLQ 2233
Db 1626 WYNKATLRMPFYGELDGCSPVLMESGLIEDHQITASSASKWSYSGTWPFPLGRLNKEG 1685
Qy 2234 RSNARPOVNPKNLEWQDKTMKVTCVTTOGVKSLLTSMYKBFLLISSQCHQWTLF 2293
Db 1686 TINAWQAKYNDMPQWLQVETQVKITGIVTOGAKFLGAEMFVTSFSLQYSHDGINWHPY 1745
Qy 2294 FQNGKV--KVFQGNODSFTPVNSLDPLRLYLRHPOSQWVHCHIALRMEVLCEAQ 2348
Db 1746 TDDGVPKAFPMGNNTNNDVKNVWYPIFSRPIRVIPKSWIGSIPMRMEFLGDCDE 1802

RESULT 8
Q804X3
ID Q804X3 PRELIMINARY; PRT: 1377 AA.
AC Q804X3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Coagulation factor VIII precursor (Fragment).
GN F8.
OS Gallus gallus (Chicken).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Tuddenham C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes.";
RL Submitted (Jan-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465272; AAC33367.1; -
DR GO; GO:0005507; F:Copper ion binding; IEA.
DR GO; GO:007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR00421; FA58 C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER
SQ SEQUENCE 1377 AA; 152728 MW; 15ACFD7F7656565CD CRC64;
Query Match
Best Local Similarity 36.3%; Score 2511.5; DB 13; Length 1377;
Matches 618; Conservative 210; Mismatches 491; Indels 385; Gaps 48;
Qy 706 LWLGCNSDFRNGMTALLKVSSCDKNTGVDYEDSYEDISAYLLKNNAIERSFS--Q 763
Db 1 VWTLCNMPDFRGMHAKFTVTQCLE-GSLDEEYEDYE---BEDFLOQRFGRK 55
Qy 764 NSRHPSTQKQFNATIPENDIEK----TDPFAHRTMPKIQNYS-SSDLLMLLRQSP 817
Db 56 NKRRPCVSEQPNNVTS-PKNGTRKPALCTES--SHEALPHNITNPGSPNGTSAPFGSP 112
Qy 818 TPGLGLSLDLOEAKYE-----TFSDDPSPGAIDSNNSLSENTHFRPOLH 861
Db 113 HPDVTSSLPETTYDYFVSYESFLADEEELSKTISQDQAGALPSRK-----H 160
Qy 862 HSDDMVFTPESGQLRLNEKLGTTAATKCLKDFKVSSTNNLSTIPSDNLAAGTDNTS 921
Db 161 ISGEVGTVSRRELQSKRPAPEDAMVRK-----VTNVL 195
Qy 922 SLGPPSPMPVHYDSQLDPTTLFGKKSSPLTSGGFLSLEENNDKLLS-GLMNSQSSWG 980
Db 196 EVQEPK-----KAAVVQAGGTLMLETTQKPMATYDLWSTVFAAG 238
Qy 981 KVSSTESGLFKGKAHGPALLTKGNALPKVISLILKINKTNNATNKRTHIDGFSLL 1040
Db 239 K-----GFL-----QCTRSSFQDDGP--- 254
Qy 1041 IENSPSVWQNLDESDFEKKVTEL-IHDRMLMDKNATALRLNHSNKTSSKNWVQOK 1099
Db 255 -EHS-----LGLQVTSSEGADVPLNLHRESRENTQPTLGSNNSFTTDS----- 300
Qy 1100 KEGPIPPDAQNPDMSPFKMLFLPESARWIQRTGKNSLNSGQSPKQIVSLGPEKSVEG 1159
Db 301 ---PLGPSARTEDIG-----TSQSHSVESNRSSSELDKLNKRPHKVVSGFNGSLGG 350
Qy 1160 QNPLSEKNVVVGKGFETVDGLKENVFPSSRNLFNTNLNHNHNTNHOEKKIQSE--- 1216
Db 351 ENI-----SFSDLQRVQEQILTDMRNSLPANGSMBEAKGTFAHSDNLEPSRYL 398
Qy 1217 -IEKKEILLQENVLQIETVTGTFKFMKNLFLSTFQNVGSGDYGAYAPVLQDFSLND 1275
Db 399 PTERDELILEAV----FODATATKOLPDTDSLVLPQSNL-----434
Qy 1276 STNRTKKTTHAFSKKGEENLEGLGNQTKQIVEKYACTTRISPTSQQNFVQTSKRAK 1335
```

```

Db 435 -----VANDTRQF-----PNALSMGVGSPRQARS 460
QY 1336 -QFR-----LPLETELEKRIIWDSTOWSKMKHLTSTLTQIDYNEKEGALTQS 1387
Db 461 LOSRGLRHLGLPSTSGRSLSLVNGAGQ-----DRSSSLTQA 502
QY 1388 PLSCLTASHIPQANRPLPIAKVSSPSPRIPTYLTRVLFDQNSHLPAAS-VRKXDSG 1446
Db 503 -----APQSAVMAASSAL-----QDNGNGTDVASNDPVSFG 536
QY 1447 VQESSHFLOGAKNNLSLAITLMTWGQORVSGLSATSNTYTKXVENVPLKPLPK 1506
Db 537 AAGRAVGLQ-----SPALAAMQ-PGGGAAMGALWGQAQGRSQLEBETNAVQPDRER 588
QY 1507 TSGKVELLPKHVHIYQKOLFPTTENGSPGLD--LVEGSLLOCTEGALKWNEANRPGVP 1564
Db 589 TW-----FSQHLAEANATENSVPSTSG-----OKPDEIP 618
QY 1565 FLRVATSSAKTPSKLLDPLAWNDHYGTQIFKEBWKQEKSPKTAFFKDKTILSLNACE 1624
Db 619 -TKVAPKENTSLP--LSSPAF--NF-----STVEXPDK----- 646
QY 1625 SNHAIANEQ-----NKPBEIVTWAKQGRTERICSQNPVPLKR-----HOREITRTL- 1674
Db 647 ---HVOASSDQGVVLGRKQVPAETGVKEEPEDEGSGSTVGKSHAAHGDNLNNETLS 703
QY 1675 --CSDQEEIDYDDTIIVEMKKEPFDIVDEENOSPRSFOKTRHYFTAAVERLWDYCMS 1731
Db 704 GLSNSESNITDYDYSITE---QFDIYGENEH--DPRSFOQIRIQYFTAAVEVITWYGNQ 759
QY 1732 SSPHLR-----NRAQSGSVPOFKKVVQFEFTDGSFTQPLVYRGEHLNHLGLGPVIRAEVE 1787
Db 760 RPQFLKATDPTSGERKPFQYRKVIRFEEYLDGSGFTQPMNGBELDEHLGILGPVIRAEIE 819
QY 1788 DNIMVITRNQASRPYSYSSLSISYEBQROQAEPKPKVFNENYKTYFWKVOHNMATKD 1847
Db 820 DTIMVYKFNLASRPFSSHQGVEE--LHGGEA---VQGEVREYSWKVLSQVMAFTMQ 873
QY 1848 EFCDKAWAYSDVDLEKDVHSGLGLPLLVCHTNTLNPAGRGQVVOBFALFFFTFDFTKS 1907
Db 874 EFDCKAWAYLSNVLDKXDLHSLGLPLLICRRGVLSNAPFKQLAVQBFSLFTFDFTKS 933
QY 1908 WYFTENNERCRAPCNTQMEDPTKENYRFAHNGYIMDTLPGLVMAQDQRIWYLLSMG 1967
Db 934 WYFENNERNCRPPCRICLQNDPDRFERNESFAHNGYVGDITLPLGLVAQQQRIWHELLNMG 993
QY 1968 SNEHISIHESGHVTVRKKEEYKALYNLYPGYFETVEMLPKAGIWRVECLIGELHA 2027
Db 994 STEDIHSHVHFGQMFSAITSOEYRMGVNLYPGVFTGVEMQPSAGIWOVECKYGEHLQA 1053
QY 2028 GMSLFLVYSNKKCOTPLGMASGH-RDFOITASGOYGQWAPKALRIHYSGSINAWSTKEPF 2087
Db 1054 GMSALFLVYNDCCNALGMASGYADSOITASGOEGWPAPKALRLHNSGSINAWSTGESN 1113
QY 2088 SWIKVDLLAPMIHGIKTQARQFSSYISQFTIMYSLDGKKWQTVRGNSGTGLMVFFG 2147
Db 1114 PWIQVDLLHLKIIHGIKTQARQKLSLYVSQFTVYFVSLHGQRWKYKGNNTTSQMMFFA 1173
QY 2148 NVDSSGIKHINFPPIIARYIRLHPTVYSIRSTIRMELMGCDLNSCMPLGMSKALSDA 2207
Db 1174 NVDATVKNRFPPIIARYIRINPTSDSIRTIURMELIGCDLNSCMPLGMSKGIPODQ 1233
QY 2208 QITASSYFTNNFATWSPSKARLHLQGRSNARWPOVNPKEMLQVDFOKTMKVTVGTQGV 2267
Db 1234 RISASSYSSNIFSSWSQARLNLQGRINARWPTNSPSEMLQVDFOKTMKVTVIITQGA 1293
QY 2268 KSLTTSYVKEFLSSODGHQWTLFPQNGKVKVQGNQDSFTPVVASLPPPLTRYLRI 2327
Db 1294 KAVFTHMVKEFVAVSSNQDGVHWSRVLHNGKEKIFRANRDYTSVLSLPPPLFARYVR- 1353
QY 2328 HPQSWVHQIALRMBVLGCEAODLY 2351

```

```

Db 1354 HPRHWHNHIALRIEFLGCDTQOEY 1377

RESULT 9
Q7SZNO PRELIMINARY; PRT; 1460 AA.
AC Q7SZNO;
DT 01-OCT-2003 (TRMBLrel. 25, Created)
DT 01-OCT-2003 (TRMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
DB Pseutarin C precursor.
OS Pseudonaja textilis (Eastern brown snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Acanthophiinae; Pseudonaja.
CX NCBI_FaxID=8673;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=22781770; PubMed=12730119;
RA Rao V.S., Swarup S., Kini R.M.;
RT "the nonenzymatic subunit of pseutarin C, a prothrombin activator from
RT eastern brown snake (Pseudonaja textilis) venom, shows structural
RT similarity to mammalian coagulation factor V.";
RL Blood 102:1347-1354(2003).
DR EMBL: AY168281; AAC38805.1; -.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 1 10 Potential.
Query Match 18.8%; Score 2340; D3 13; Length 1460;
Best Local Similarity 25.68; Pred. No. 1.2e-130;
Matches 601; Conservative 287; Mismatches 511; Indels 952; Gaps 44;

QY 22 RRYLGAVELSWMY--MQSDLGELPVDARPPRPVKSPFPTSVVYKTLFVFTDHLFN 79
Db 34 REYHIAAQLEDWDYNPQPELSRLS-----ESDLTFKKIVREY-ELDFX 77
QY 80 IAKERPMMGLGPTQIAEYVDVTWITLKNWASHPYSLHAGVSYMKASGAEVDQTSQ 139
Db 78 QSEPRDALSGLLGPTLRGEVGDLSLIYFRNFATQPSIHQPSAVYNNKWSGSSYSDGTS 137
QY 140 REKEDDKVFPGGSHYVYQVLKNGPMASDPLCLTYSLSHVDLVKDLNSGLIGALIVCR 199
Db 138 VERLDDAVPQGFQKVVNNITAEIGPKKADPPCLTYVYSHVAVMVRDFNSLIGALLICK 197
QY 200 EGSIAKEKTQTL--HKFILLPAVFDGKSWHSETKNSLMODRDAASARAPKMHVNGYV 257
Db 198 EGSINANGSQKFTNREYVLMFSVFDSEKNWY--RKPSL-----QYTINGFA 241
QY 258 NRSLPGLIGCHRSVYVWHVIGMGTTPVHSIFLEGHTFLVRNHROASLEISPIFTLTAQT 317
Db 242 NGTLPDQVQACAYGHISHLIGMSSSEIFSVHNGQTLQNHVYKVTINLVGGASVTAD 302
QY 318 LMDLGOFLLSCHISGHQHDGMEAYVYKVDSCPEEPQLRMKNKEAEHDDDLTDSMDVY 377
Db 302 SVSRTGKWLISLVAHXHQAGMYGLYNIKDCGNPDULTRK----- 341
QY 378 RPDNDNSPSIQTRSVAKKXPKTWVYHIAAEEDWDYAPLVLPADDRSKYSQVLLNNGPQR 437
Db 342 -----LSPRELMT-----KNWEYFIAABETWDYAPBIPSVDPRRYKAQYDLNFSNF 389
QY 438 IGRKYKVRVMAYTDTF-KTREAI--QHSIGLGPLLYGVBGDTLLIIFKNOASRPVNIY 495
Db 390 IGKYYKAVPROVEDGNFTKPTVAVPKRGILGPNVIAKVRDTVIVFKNLASRYSIY 449
QY 496 PHGITDVR-----PLYGRRLPFGYKHLKOPFLPGEIFKYKWTVTVDGKTSDFCLTRY 551
Db 450 VHGVSVSKDAEGAIYPSDPKENITGCK--AVEPGQVYVTKWTVLTDDEPTVKDECIITKL 507
QY 552 YSSFVAMERDLASGLIGPLLIICYKESVDQRCNQIMSKDKNVLFSVFDENRSLYLTENIQ 611
Db 508 YHSAVDMTRDIASGLIGPLLVCKEKALS VKGVQNKADVEQHAVFAVFDENKSWYLEDNK 567

```



```

QY 1956 NLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYGNKQCTPLGMAHGHIRDPQ 2055
Db 1806 PLLP
QY 2056 ITASQYQOWAPKLARL--HYSGSIN-AWSTKEPFSWKVDLLAPWLIHGIKTQARQKF 2112
Db 1827 IKAS-EY-----LSCLGCHARGLDTPVWYSNP-DKIPRSFLIIRLL-----PKKY 1871
QY 2113 SSLYSIQIILYVSLDGKKQWTRGNSGTGLMVFFGNVDSGKHNIFNPPIIARIIRLHP 2172
Db 1872 SNAESLSVILV-----QTLISLFLVLYFAGNSDASTIKENRFPPIVARIRHP 1923
QY 2173 THYSIRSTRILMELMCDLNSCMPILGMRSKALSDAQITASSFTNMFAT-NSPSKARLHL 2231
Db 1924 TKSYNRPILRLLELCEVNGGCTPLGLEDRIGTQNKQITASSFKKSWGSGWYEPSPARLNA 1983
QY 2232 QORSNAWRPQVNPKEWLQVDPOKTMKVTVGVTQGVKSLTSMYVKEFLISSQDGHWT 2291
Db 1984 QQRVNAWQAKANNKQWLQIDLLKIKKVTAVTQCKSLSEMYVKS:SIILYDQGVSWK 2043
QY 2292 LFFQNGKV--KVFQNGQSFVNVNSLDPPLTRVLRHPOSVWHQIALRMEVLGCEAQD 2349
Db 2044 PYRQKSMVDKIFEGNSNTKGMKNFNPPIISRFIRIIPKTNQSIARLELFC---D 2100
QY 2350 LY 2351
Db 2101 IY 2102

RESULT 11
Q8BQ43 PRELIMINARY; PRT; 355 AA.
ID Q8BQ43
AC Q8BQ43;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Coagulation factor VIII (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK051579; BAC34681.1; -.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR Pfam; PF00394; Cu-oxidase; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 355
SQ SEQUENCE 355 AA; 40548 MW; 2D3C984EFA43F284 CRC64;

Query Match 12.9%; Score 1595.5; DB 11; Length 355;
Best Local Similarity 84.2%; Pred. No. 3.8e-87;
Matches 298; Conservative 22; Mismatches 33; Indels 1; Gaps 1;

QY 1 MOLESTCFELCLLRCPFSATRYILGAVELSDYMQSD-LGELPVDARPPVPKSPFF 59
Db 1 MQALFACFFLSLNFPCSSAIRRYILGAVELSNWYIQSDLSVLHTDSRFLPRMSTSFPP 60
QY 60 NTSVWYKTLDFVETDHLFNIAKRPWMLGLGPTTCAEYVYVFWITLKNMASHPVSLSHA 119
Db 61 NTSIMYKTVFVYKQLFNIAKRPWMLGLGPTTWEVDHTVITLKNMASHPVSLSHA 120
QY 120 VGVSYWKASEGAYDDQTSQREKDDKVPFGSGHTYVWVLKENGPMASDPCLCLTYSLS 179
Db 179 VGVSYWKASEGAYDDQTSQREKDDKVPFGSGHTYVWVLKENGPMASDPCLCLTYSLS 179

```

```

Db 121 VGVSYWKASEGAYDDQTSQREKDDKVPFGSGHTYVWVLKENGPMASDPCLCLTYSLS 180
QY 180 HVDLVKDINSGLIGALLVCREGSLAKEYQTLHKFILLFAVDFEGKSMHSETKNSLMQDR 239
Db 181 HVDLVKDINSGLIGALLVCREGSLAKEYQTLHKFILLFAVDFEGKSMHSETKNSLMQDR 240
QY 240 DAASARAPKMTVNGYVNRSLPGILGCHRSVYWHVIGMGTTPVHIFLECHTFLVRN 299
Db 241 DSASARDPKNMTVNGYVNRSLPGILGCHRSVYWHVIGMGTTPVHIFLECHTFLVRN 300
QY 300 HQASLEISPIITFLTAQTLMDLGFLLSCHISSHOHDGMEAYVKVDSCEPEPQ 353
Db 301 HQASLEISPIITFLTAQTLMDLGFLLSCHISSHOHDGMEAYVKVDSCEPEPQ 354

RESULT 12
Q804X4 PRELIMINARY; PRT; 745 AA.
ID Q804X4
AC Q804X4;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Coagulation factor V precursor (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RL coagulation genes from Gallus gallus and Fugu rubripes.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465271; AAC33366.1; -.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FAS8 C.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS50022; FAS8C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 745
SQ SEQUENCE 745 AA; 85626 MW; A264587A348C29EE CRC64;

Query Match 11.7%; Score 1453; DB 13; Length 745;
Best Local Similarity 39.7%; Pred. No. 4e-78;
Matches 299; Conservative 143; Mismatches 258; Indels 54; Gaps 15;

QY 1626 NHAIAAINGONKPEIEVTWAKQGRTERLCSQNPVVL--KRHOR-----EITFTLQ 1675
Db 9 NHCCVAIL--PQSHVPSFATENCQGYTQQL---KPSIVIGLPHENGNYEYTSGEYVETDS 63
QY 1676 SQDEE---IDYDDTISVENMKEDFDIYDRDNQSPSPFK-----KTHYFIAAVERL 1725
Db 64 GDEYEVYVSPDDPYMTDFK-----LVNVEGRNDNIAEHYLRSGNRERRYIAAKYVC 117
QY 1726 WDF-GMSSPHVRNRAQSGSVQPKVQVQFTDGSFTQPLRYRGLNEHLGLGVIRA 1784
Db 118 WNYAGLKSTWMDYKICKDQST--KKVIFQYTDSTFTTVQDEGEYRELGLGVIRA 174
QY 1785 EVEDNIMVTRNQARPPYSPYSSLSIYE-----EDQQAEPKPKVKNETKTYFW 1836
Db 175 EYNDVILVHKLASRPYSLHAHGLFYEKSGISYDDESSDWFKBDKVPQNSSVIYW 234

```

```
QY 1837 KVQHMAPTKDEPDKAWAFSDVLEKDVHSHGILGLPILVCHNNTLNPAHGRQVTVQEF 1896
Db 235 YANRSQPLQGAACRSWVYSDVNLKSDTHSLGILGLPILCEKTSFSDNNRINURDFF 294
QY 1897 LFTIFDETKSWFTENWERNCRAPCNIOWEDPTFKENYRPHALNGYIMDTLGLUWAQD 1956
Db 295 LLEWVPEERSWTF---DKRSKXCTEKTQE--MQOCHRFYANG-ITVNLQGLRMYEG 347
QY 1957 ORTRWILLMSGNENIHSIHFSGHVFTVRKKEYSKALYNLYPGVFTVEMLPSKAGIWR 2016
Db 348 ELVRWHLNMGVDPKHVHFHGTFTFTELEAKYQLGTITLLPGSTQTVEMKQRPQGWL 407
QY 2017 VECLICEHLHAGMSTLFLVYNSKCPGLNAGSHIRDPQTASGGQGWAPKARLHYSG 2076
Db 408 LDEVEGEYQAGQASVLYLEKGRIPMGWASGVILDSQINAGSHHIDYEMPKLARLNSG 467
QY 2077 SINAWST--KEPFSWIKVILLAPMIHIGIKTQARQKFSLSYLSQILYSLDGKWKQTY 2134
Db 468 TVNAWSTWTCEDLPWLOVDFQORVILGTGTQAKHFLKSPYQKLFIVYSTDKRWNI 527
QY 2135 RGNSTGLVAFGNVSSGIGIKHFNPPIIARVIRLHPHYSIRSTRMLMGLMCDLNSGS 2194
Db 528 RGDSSPAEKIFEGNSDAYGIKENIIDPPIIARVIRVVPKQAVNEPTLRMELLCGEVDACS 587
QY 2195 MPLGMSKALSDAQITASSVFTWVATWSPSKARLHLQGRSNARVQVANKZWLQVDFQ 2254
Db 588 LPLGMEGKIKNTQITASSVFTWVATWSPSKARLHLQGRSNARVQVANKZWLQVDFQ 647
QY 2255 KTWKVTGVTQGVKSILTSYKFEPLISSQDGHQWTLFFQNGK--VKVFGNQDSFTTV 2312
Db 648 TIKKATAIATQGVKSVTFNFVKTYVILYSNQSEWKSYTESSSVAKVFGNEDSRGHV 707
QY 2313 VNSLOPPLTRVLRHPQSWFQIALRMEVLGCE 2346
Db 708 KHFFNPPLSRFTRIVPKTWYANGIALRVELFGCD 741

RESULT 13
Q80Y80
ID Q80Y80 PRELIMINARY; PRT; 1156 AA.
AC Q80Y80;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hephaestin (Heph protein).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton K., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.C., Du X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
```

```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049237; AAH54442.1; --
DR EMBL; BC054442; AAH54442.1; --
GO; GO:0005507; F: copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR002355; Cu-ox_copper_BS.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1156 AA; 129567 MW; 5F11033599DC1874 CRC64;

Query Match 11.2%; Score 1389; DB 11; Length 1156;
Best Local Similarity 20.0%; Pred. No. 5.4e-74;
Matches 412; Conservative 180; Mismatches 413; Indels 1052; Gaps 26;

QY 20 ATRRYYLGAVALSWDMQSDGLGELPVDARFPFPRVPKSP-----PNTSVVYKTLFVEPTD 75
Db 24 ATRNYYLGLQDMQWYAKGRNVITQNLINDTVASSFLKSGKNRIGSYKTYVKEYSD 83
QY 76 HLFNTAKPRPPMGLGFTIQAQVYDTVVITLKNMASHPVSLHVGVSXKASGAEYDD 135
Db 84 GYTBETIAKPAWLGFLGQLQAQVGVSVILHLKNFASRPTIHPHGVFEKDSLSLPD 143
QY 136 QTSQREKDDKVPFGSGSHYVWVLKENGPMASDPCLTYSYLSHVDLVKDLNSGLIGAL 195
Db 144 GSGYLKADDSVPPGSGSHVYNSIPESHAPTEADPACLTWIYHSHVDAPDRIATGLIGFL 203
QY 196 LVCREGSL-----AKETQTLKFLLEPAVEDEGKSWHE-----TKNSLQMDRAAASAR 245
Db 204 TCKGKGLDGNSPFQKQVDNFFLLFVSDENLSWHLDDNIAITYCSPASVSKDGAQF 263
QY 246 AWPQKHTVNGYVNSLPLGLIGHCRKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASL 305
Db 264 DSNRMHAINGFVGNLPGLSLMSCAQKVAWHLFGMNEIDVHTAFHGMQLSIRGHHTVA 323
QY 306 BISPTFTEAQTLLMDLQFLLSCHISQHDGDMZAYVKVDCSPBPOLRMKNNEAEADY 365
Db 324 NIFFATFVAENKVFQKSGTFLSCVNSHLRSGMQAFYKYVDCSCMDPPV----- 372
QY 366 DDLTDSMDVYVRFDDDDNSPSFIQIRSVAKKEPKTWVHYIAAEBEDWDYAPLVLAPDSS 425
Db 373 -DQLTG-----KVRQYFQAHIQWDYGP--IGYDORT 402
QY 426 YKS-----QYLNNGPQIRGRYKVRVMAYDTFTKTRTAIQHES--GILGPLLYG 474
Db 403 GKSLREPGSGPKYFQKSSSSRIGTYKVRVEAFQDETQFQSRVHQEETHLGLGPVIRA 462
QY 475 EVGDTLLIFKNOASRPVNIYPHGITDVRPLYSRLPKGVKH---LKDPFPLPGEIIFYK 531
Db 463 EVGDTIQVFNRSQFSPQPHGVFVKEKSGTGYNDGTSHPKVAKSF-----EKVTYY 517
QY 532 WTVTVEDGPTKSDPRCLTRYYSFVNMRDLASGLIGLILICYKESVDQRGNQIMSKSN 591
Db 518 WTVVPPHAGTAQDAFLTWVMYFSAADPTDNTNSGLVGLVCKAGALGADQKGVDFKEF 577
QY 592 VILFSPDEVRNRYLTENIORFLENPAG-----VOLEDPF--FQASNTMHSINGYVFDL- 644
Db 578 FLLETFVDENESWYSAN-----QAAGMLDSRLSSDVEGFQDSNRMHAINGFLFSLNP 631
QY 645 QLSVCLHEVAYWYILSGAOTDPLSVFSGVTFKHVMYEDTLTFPFSGETVPMSEN 704
Db 632 RLDMCKGDTVANHLGLGTETDVEGVNFBGNTVOLQGRKGAVMFLPHTTVAIMQDPNP 691
```

```
QY 705 GLWILCHNSDFRNGMTALLKVSSCDKNTGDTYYEDSYEDISAYLLSKNNAIPERSFQN 764
Db 692 GIPEIYCQAGSHREBGMQAIYNSQC-----SSHQD-----SPRQHYQA 730
QY 765 SRHPSTRQKQFNATTIPENDIEKTDWFAHRTWPKXIQNVSSDDLMLLRQSTPGLSL 824
Db 731 SR-----732
QY 825 SDLOEKYETFDSPGPAIDSNNSISEMTHFRQLHSHGDMVFTPEGQLRLNEKIGT 884
Db 733 -----732
QY 885 TAATELKCLDFKVSSISNNLISIPSDNLAAGTNDTSSGLPPSPMVFHYQSOLDTTLFGKK 944
Db 733 -----VYV-----735
QY 945 SGPLTBSGGPLSLSEENNDSKLLSGLMNSQESSWGKNVSSYTESGRLFKGKAHGPALLT 1004
Db 736 -----IMAEELW-----743
QY 1005 KDNALPKVISILKTKNTSNNATNRKTHIDGFSLLIENSPPVWQNILESDETEFKKVTEL 1064
Db 744 -----DYCPDRSWELE-----WNTSEKD-----762
QY 1065 IHDRMLMDKNATALRLNHSNKTSSKNMEMVQOKKEGIPDDAQNPDMSPFFKMLPLPES 1124
Db 763 -----762
QY 1125 ARWIORTHGKNSLNSQGPSPKQLVSLGPEKSVSQNLSEKNVVGKGBTKDVGLE 1184
Db 763 -----762
QY 1185 MYFPSSRNLFJNLNHNHNTHNOEKKTOEIEKETLIOENVVLPQIHTVTGTNPFMK 1244
Db 763 -----SYGHVFLSKND-----773
QY 1245 NLFLLSTRNGVEGSDGAPVLPQDFRSLNDSTNRKTHAHSKKGEBNLEGLGNQTK 1304
Db 774 -----773
QY 1305 QIVEKYACTRISPNSTSQNFVTRQSKRALKQPLPLEETELEKRIIVDDTSTQSKNMK 1364
Db 774 -----773
QY 1365 HLTPSTLTQIDYNEKEKGAITOSPUSDCLTRSHSIPOANRSPPIAKVSSPSIRPIYLT 1424
Db 774 -----773
QY 1425 RVLFDNSSLPLPAASYRKDQSGQESSHFLQAKKNNLSLITILEMTGDQREVSGLGTS 1484
Db 774 -----GLLGS-----778
QY 1485 ATNSVTYKKVENTVLPKPLPNTSGKVELLPKVHYQKDLFPTETSGSPGHLDLVEGSL 1544
Db 779 -----778
QY 1545 LQTEGAIKWNENRPGKVPFLRVATESAKTPSKLLDPLADWNHYGTIPKEWKVSQEK 1604
Db 779 -----778
QY 1605 SPEKTAFKKDTILSLNACSNHAIAINEGQNKFEIEVTWAKQORTBELCSQPPVLKR 1664
Db 779 -----778
QY 1665 FQREITRITLQSDQSEIDYDDTISVEMKEDFDIYDEENGSPRSFQKKTRHYPIAAVER 1724
Db 779 -----778
QY 1725 LWDYGMSSSPHLNRAGSGSVOPKVVQBFQDTSPTQPLYGELNEHGLJLGPYIRA 1784
Db 779 -----RYKKAVPREYTDGTFRIIPRPSGPEHGLGILPLIRG 815
QY 1785 EVEDNIMVTPRNQASRPYSFYS-SLISVEEDQROCAEPRKNFVKPNETKITFWKQVQHMA 1843
```

```
Db 816 EVGDIITVVKKASRPYSIHAGVLESNTGGQARZ-----PGEVITYQWNIERSG 868
QY 1844 PKDEPCKAWAYFSDVDLEKQVHSLGLOPLVCHTNTLNPAHGRQVTVQEFALFTTFID 1903
Db 869 PGFSDSACSWIYISAVDPIKDWYSLGVPLVICRNGILEPNGGRNDMDREPALFLIFD 928
QY 1904 ETKSWYFTENMER-NCRAPCNTQMEDPTKENVRPHAINGYIMDTPLGLVMAQDQIRWY 1962
Db 929 ENQSWLTKENIZATYGPQESSHVNLKDATFLEGNKHAINGKLYANLRGHTVYQGERVAVY 988
QY 1963 LLSMSNENIHSIHPFSGHVFTVRKKEEYKVALYNLYPGVFETVEMLPKSKAGIWRVECLIG 2022
Db 989 MLAMGQDITDIIHVHFAESFLYQNGQSYRADVDVLPFGTFEVVEMVASNPGIWLMECHVT 1048
QY 2023 EHLHAGMSTLFLVYNNK 2039
Db 1049 DHVHAGMETIFTVLSHE 1065
```

RESULT 14

```
Q9Z0Z4
ID Q9Z0Z4 PRELIMINARY; PRT; 1157 AA.
AC Q9Z0Z4;
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Hephaestin.
GN HEPH.
OS Mus musculus (Mouse).
OC Zukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=99140771; PubMed=9888272;
RA Vulpe C.D., Kuo Y.M., Murphy T.L., Cowley L., Askwith C., Libina N.,
RA Gitschier J., Anderson G.J.;
RT "Hephaestin, a ceruloplasmin homologue implicated in intestinal iron
RT transport, is defective in the sla mouse.";
RL Nat. Genet. 21:195-199(1999).
DR EMBL; AF082567; AAD16035.1; -.
DR HSSP; P00450; 1KCM.
DR MGD; MGI:1332240; Heph.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR006972; Cupredoxin.
DR InterPro; IPR002355; Cu ox copper_BS.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1157 AA; 129682 MW; 4119BIDA2BFE2CFD CRC64;
```

```
Query Match 11.2%; Score 1388; DB 11; Length 1157;
Best Local Similarity 20.0%; Pred. No. 6.2e-74;
Matches 412; Conservative 179; Mismatches 414; Indels 1052; Gaps 26;

QY 20 ATRYVYLGAVELSDWYMQSDLGELPVDARPPRPVKSF----PNTSVVYKKTFLVBFTD 75
Db 24 AIRNYLIGIQDMQNYAPKGRNVITNQTNNDTVASSFLKSGKNRIGSSYKKTVKEYSD 83
QY 76 HLFNIAKPRPPMGLTGTTQAEVYDTVITLKNWASHPYSLHAVGVSYWKASGAYDD 135
Db 84 GTYTEETAKPAWLGFLGLQAEVGVILLHLKNTASRPYTIHPGVYFKEQSGSYPD 143
QY 136 QTSOREKDDKVPFGSGSHYVQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLICAL 195
Db 144 GSSGLKADDSVPFGSGSHVYNWSIPESHAPTEADPACLTWIYSHVDAPDIATGLIGPL 203
QY 196 LVCHREGSL---AKEKTQTLHKFILLFAVFDEGKSWHSE-----TKNSLMQDRDAASAR 245
Db 204 ITCRGTLDGNSPPQKRDVDHNFLLFSVIDENLSWHLDDNIATYCSDPASVDKEDGAFQ 263
```


QY	1957	QIRWYLSMGSNENIHSIHFSGHVFTVEKKKEYKVALYNLYPGVZETVEMLPKAGIWE	2016
Db	983	ERVAVMYLAYSQDTHHTVHFHAESEFLYQNGHSYRADVDLFFGTFFVEMVANSNPGAWL	1042
QY	2017	VECLIGERLHAGMSTLFLYYSNK	2039
Db	1043	MECHVTDHVEAGMETFTVLSHE	1065

Search completed: April 13, 2004, 14:07:51
Job time : 103.5 secs

GenCore version: 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 22.5 Seconds
(without alignments)
5394.340 Million cell updates/sec

Title: NP000123-328
Perfect score: 12416
Sequence: 1 MQIELSTCFLLRFCFSFA.....WVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pdp:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pdp:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pdp:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pdp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pdp:*
6: /cgn2_6/ptodata/2/iaa/backfiles.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12410	100.0	2351	4	US-10-133-907-4
2	12404	99.9	2351	6	Sequence 4, Appli Patent No. 5171844
3	12401	99.9	2351	1	Sequence 2, Appli Patent No. 5171844
4	12397	99.8	2351	1	Sequence 2, Appli Patent No. 5422260
5	12388	99.8	2351	6	Sequence 4, Appli Patent No. 5422260
6	12293	99.0	2332	1	Sequence 2, Appli Patent No. 5422260
7	12293	99.0	2332	1	Sequence 4, Appli Patent No. 5422260
8	12293	99.0	2332	1	Sequence 2, Appli Patent No. 5422260
9	12293	99.0	2332	1	Sequence 4, Appli Patent No. 5422260
10	12293	99.0	2332	1	Sequence 2, Appli Patent No. 5422260
11	12293	99.0	2332	3	Sequence 2, Appli Patent No. 5422260
12	12293	99.0	2332	4	Sequence 2, Appli Patent No. 5422260
13	12293	99.0	2332	5	Sequence 2, Appli Patent No. 5422260
14	12293	99.0	2332	5	Sequence 2, Appli Patent No. 5422260
15	12293	99.0	2332	5	Sequence 2, Appli Patent No. 5422260
16	12286	99.0	2332	3	Sequence 2, Appli Patent No. 5422260
17	12243	98.6	2332	3	Sequence 2, Appli Patent No. 5422260
18	9485	76.4	2343	3	Sequence 2, Appli Patent No. 5422260
19	8827	71.1	2319	1	Sequence 2, Appli Patent No. 5422260
20	8827	71.1	2319	1	Sequence 2, Appli Patent No. 5422260
21	8827	71.1	2319	2	Sequence 2, Appli Patent No. 5422260
22	8827	71.1	2319	3	Sequence 2, Appli Patent No. 5422260
23	8827	71.1	2319	4	Sequence 2, Appli Patent No. 5422260
24	8827	71.1	2319	4	Sequence 2, Appli Patent No. 5422260
25	8827	71.1	2319	5	Sequence 2, Appli Patent No. 5422260
26	8722	70.2	2304	3	Sequence 2, Appli Patent No. 5422260
27	8349	67.2	1661	2	Sequence 2, Appli Patent No. 5422260

28	8349	67.2	1661	2	US-08-558-107-2	Sequence 2, Appli
29	8349	67.2	1661	3	US-09-243-539-2	Sequence 2, Appli
30	8068	65.0	2133	2	US-08-670-707A-37	Sequence 37, Appli
31	8068	65.0	2133	3	US-09-037-601-37	Sequence 37, Appli
32	8068	65.0	2133	4	US-09-315-179-37	Sequence 37, Appli
33	8068	65.0	2133	4	US-09-523-656-30	Sequence 30, Appli
34	7951.5	64.0	2115	3	US-09-324-867-5	Sequence 5, Appli
35	7392	59.5	1471	1	US-08-683-8393-3	Sequence 3, Appli
36	7219	58.1	1438	1	US-09-203-916-1	Sequence 1, Appli
37	6216	50.1	1467	4	US-09-523-656-38	Sequence 38, Appli
38	6114	49.2	1443	2	US-08-670-707A-39	Sequence 39, Appli
39	6114	49.2	1443	3	US-09-037-601-39	Sequence 39, Appli
40	6114	49.2	1443	4	US-09-315-179-39	Sequence 39, Appli
41	2774.5	22.3	2183	3	US-08-746-111-5	Sequence 5, Appli
42	2436.5	19.6	541	1	US-08-121-202-4	Sequence 4, Appli
43	2258	18.2	435	3	US-08-448-722A-5	Sequence 5, Appli
44	2258	18.2	435	3	US-08-189-309B-5	Sequence 5, Appli
45	2190.5	17.6	868	1	US-07-864-004B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-133-907-4
; Sequence 4, Application US/101333907
; Patent No. 6677369
; GENERAL INFORMATION:

; APPLICANT: Chien, Kenneth R
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10/133.907
; PRIOR FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-907-4

Not 9150

Query Match	100.0%;	Score	12410;	DB	4;	Length	2351;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	2350;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						
Qy	1	MQIELSTCFLLRFCFS	ATRRYYLGA	VELSWD	YMO	SDLGEL	PVDARFP
Db	1	MQIELSTCFLLRFCFS	ATRRYYLGA	VELSWD	YMO	SDLGEL	PVDARFP
Qy	61	TSVYKKTLLFVEPTDHL	FNIAKPRPP	MMGLIPT	QAEVYD	TVITL	KNMASHPVS
Db	61	TSVYKKTLLFVEPTDHL	FNIAKPRPP	MMGLIPT	QAEVYD	TVITL	KNMASHPVS
Qy	121	GVSWKASEGAEYDDQTS	QREKDDKVP	PGSGSHY	WQVLK	ENGPMAS	DELCLTYS
Db	121	GVSWKASEGAEYDDQTS	QREKDDKVP	PGSGSHY	WQVLK	ENGPMAS	DELCLTYS
Qy	181	VDLVKLNGLIGALLVC	REGSLAKE	KTOTLH	FILLFA	VFDEK	SHSTKNS
Db	181	VDLVKLNGLIGALLVC	REGSLAKE	KTOTLH	FILLFA	VFDEK	SHSTKNS
Qy	241	AASARAWPKMHTYNG	YVNRSLPGL	IGCHRSV	YWHVIG	MTTPEV	HSIFLE
Db	241	AASARAWPKMHTYNG	YVNRSLPGL	IGCHRSV	YWHVIG	MTTPEV	HSIFLE
Qy	301	QASLHISPTITFTAC	TLLMDLQ	QFLLSCH	ISSHQH	DGMAY	VKVDC
Db	301	QASLHISPTITFTAC	TLLMDLQ	QFLLSCH	ISSHQH	DGMAY	VKVDC

361 EAEDYDDDLTDSBMDVVRFDKSPSPFIQIRSVAKKHPKTVWVHIAAEEDBDWVAPLVL 420
Db
361 EAEDYDDDLTDSBMDVVRFDKSPSPFIQIRSVAKKHPKTVWVHIAAEEDBDWVAPLVL 420
QY
421 PDRRSYKSOYVNNNGPQRIGSKYKVKVPMAYTDTFTKTREAIQHESGILGPLLXGEVGDIL 480
Db
421 PDRRSYKSOYVNNNGPQRIGSKYKVKVPMAYTDTFTKTREAIQHESGILGPLLXGEVGDIL 480
QY
481 LIIFKQASRPVNIYPHGIDTVRPLSRRLPKGVGHLKDPPLIPGBIFKVKVTVTVEDGP 540
Db
481 LIIFKQASRPVNIYPHGIDTVRPLSRRLPKGVGHLKDPPLIPGBIFKVKVTVTVEDGP 540
QY
541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQSGNQIMSDKRNVLFSVDFE 600
Db
541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQSGNQIMSDKRNVLFSVDFE 600
QY
601 NRSWYLITENIQRELPNAGVQLBDDPFOASNIHVSINGYVFDLSQLSVCLHEVAYWYLS 660
Db
601 NRSWYLITENIQRELPNAGVQLBDDPFOASNIHVSINGYVFDLSQLSVCLHEVAYWYLS 660
QY
661 IGAQTFPLSVFRSGYTFKHXYVEDTLTLPPPSGETVSMENPGLWILGCHNSDPRNRG 720
Db
661 IGAQTFPLSVFRSGYTFKHXYVEDTLTLPPPSGETVSMENPGLWILGCHNSDPRNRG 720
QY
721 MTALLKVSXCDKNTGDIYEDSYEDI SAYLLSKNAIEPRSFQNSRHPSTROKQFNATTI 780
Db
721 MTALLKVSXCDKNTGDIYEDSYEDI SAYLLSKNAIEPRSFQNSRHPSTROKQFNATTI 780
QY
781 PENDIEKTDWPAHRTMPKPIQNVSSDDLMLLRQSPTHGLSLSD:QBAKYETFSDDPS 840
Db
781 PENDIEKTDWPAHRTMPKPIQNVSSDDLMLLRQSPTHGLSLSD:QBAKYETFSDDPS 840
QY
841 PGAIDSNNSLSEMTHERPOLHSGDMVFTPEGLQLRLNKLGTAAATELKLKLDKVSST 900
Db
841 PGAIDSNNSLSEMTHERPOLHSGDMVFTPEGLQLRLNKLGTAAATELKLKLDKVSST 900
QY
901 SNMLISTIPSDNLAAGTNDTSSIGPPMPVHYDSQDITLFGKKSFLTSESGPLSLSE 960
Db
901 SNMLISTIPSDNLAAGTNDTSSIGPPMPVHYDSQDITLFGKKSFLTSESGPLSLSE 960
QY
961 NNDKILLESGLMNSORSSGKNVSSYTESGRLFGKRAHGPAITKONALFKVSI SLLKTN 1020
Db
961 NNDKILLESGLMNSORSSGKNVSSYTESGRLFGKRAHGPAITKONALFKVSI SLLKTN 1020
QY
1021 KTSNNSATNRKTHIDGFSILLIENSPSVWQNI LBSDTFEEKVVTPLIHDRMLXDNATLAL 1080
Db
1021 KTSNNSATNRKTHIDGFSILLIENSPSVWQNI LBSDTFEEKVVTPLIHDRMLXDNATLAL 1080
QY
1081 NMSNKTTSKKNEMVQOKKEGPIPDQAQNPDMSPFKMLFLPESARWIORTHGKNSLNG 1140
Db
1081 NMSNKTTSKKNEMVQOKKEGPIPDQAQNPDMSPFKMLFLPESARWIORTHGKNSLNG 1140
QY
1141 QGPPSPKQYLSLGPKEGVEGNFSEKNKVVGKGEFTKOVGLKEMVFPSSRNLFLLNLDN 1200
Db
1141 QGPPSPKQYLSLGPKEGVEGNFSEKNKVVGKGEFTKOVGLKEMVFPSSRNLFLLNLDN 1200
QY
1201 LHNNTNHOEKKIQE3IEKKEKTELIQENVLPQIHTVTGKNFMKNLFLSTRQNVESYD 1260
Db
1201 LHNNTNHOEKKIQE3IEKKEKTELIQENVLPQIHTVTGKNFMKNLFLSTRQNVESYD 1260
QY
1261 GAYAPVLOPFRSLNDSNTKHTAHFSPKKGEBENLEGONQTKQIVEKYACTRISPN 1320
Db
1261 GAYAPVLOPFRSLNDSNTKHTAHFSPKKGEBENLEGONQTKQIVEKYACTRISPN 1320
QY
1321 SQONFVTOQSKRALKQFRLPLEBETELEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKE 1380
Db
1321 SQONFVTOQSKRALKQFRLPLEBETELEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKE 1380
QY
1381 KGAITQSPSLDCLTRSHSIPQANRSLPIAKVSSPFSIRPIYLTRVLFQDNSSHLPAASY 1440
Db
1381 KGAITQSPSLDCLTRSHSIPQANRSLPIAKVSSPFSIRPIYLTRVLFQDNSSHLPAASY 1440
QY
1441 RKDQSVQSSSHFLQAGAKNNLSLAITLMTGDQREVSLGTSANSVTYKKVENTVLP 1500

1441 RKDQSVQSSSHFLQAGAKNNLSLAITLMTGDQREVSLGTSANSVTYKKVENTVLP 1500
QY
1501 KPDLPKTSQKVLLPKVHIYQKDLFPFTETSNQSPGHLDLVEGSLLOQTEGAIKWEANRP 1560
Db
1501 KPDLPKTSQKVLLPKVHIYQKDLFPFTETSNQSPGHLDLVEGSLLOQTEGAIKWEANRP 1560
QY
1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWKSOEKSPKTAFAKKOTILSL 1620
Db
1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWKSOEKSPKTAFAKKOTILSL 1620
QY
1621 NACESNHIAAANEGONKPEIEVTWAKQORTERLSONPVLKXHQREIRTRITLQSDOE 1680
Db
1621 NACESNHIAAANEGONKPEIEVTWAKQORTERLSONPVLKXHQREIRTRITLQSDOE 1680
QY
1681 IYDDDTISVEMKKE3PDIYDEBENQSPRSFQKTRHYFAAVERLWDYQMSSPHVLNR 1740
Db
1681 IYDDDTISVEMKKE3PDIYDEBENQSPRSFQKTRHYFAAVERLWDYQMSSPHVLNR 1740
QY
1741 AQSGSVQPKKVVFQEFDTGSGFTQPLRYGELNEHGLGPGYIRAEVEDNIMVTFRQASR 1800
Db
1741 AQSGSVQPKKVVFQEFDTGSGFTQPLRYGELNEHGLGPGYIRAEVEDNIMVTFRQASR 1800
QY
1801 PYSFYSSLSIYEBEDORQGAEPKKNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSV 1860
Db
1801 PYSFYSSLSIYEBEDORQGAEPKKNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSV 1860
QY
1861 DLEKOVHSLGILGFLVCHTNTLNPAHGRQVTVQEPALFPTI FDETKSWYFTENMERNCSA 1920
Db
1861 DLEKOVHSLGILGFLVCHTNTLNPAHGRQVTVQEPALFPTI FDETKSWYFTENMERNCSA 1920
QY
1921 PCNIQMEDTFRKENVREHAINGIIMDTPLGLVMAODQRIWYLLSNGSNENIHS:HFSGH 1980
Db
1921 PCNIQMEDTFRKENVREHAINGIIMDTPLGLVMAODQRIWYLLSNGSNENIHS:HFSGH 1980
QY
1981 VFTVRKKEBYKMALXNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Db
1981 VFTVRKKEBYKMALXNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY
2041 QTFPLGMAASHIEDFOITASQYQWAPKLARLHYGSGINAWSTKBPFSWIKVDLLAPMII 2100
Db
2041 QTFPLGMAASHIEDFOITASQYQWAPKLARLHYGSGINAWSTKBPFSWIKVDLLAPMII 2100
QY
2101 HGKIQCARQKPSLSYISOFIIMYSLDGKKWOTYRGNSTGTLMVFPNGVSSGKHNIFN 2160
Db
2101 HGKIQCARQKPSLSYISOFIIMYSLDGKKWOTYRGNSTGTLMVFPNGVSSGKHNIFN 2160
QY
2161 PPIIARYIRLHPHTHYSIRSTLMEMLMGCDLNSCMLPLGMESKAI SDAQITASSYFTNMFA 2220
Db
2161 PPIIARYIRLHPHTHYSIRSTLMEMLMGCDLNSCMLPLGMESKAI SDAQITASSYFTNMFA 2220
QY
2221 TWSPKARLHLOGRANARPOVNNPKMLQVDPKTMKVTGVTGCVKSLTSMYVKBFL 2280
Db
2221 TWSPKARLHLOGRANARPOVNNPKMLQVDPKTMKVTGVTGCVKSLTSMYVKBFL 2280
QY
2281 ISSSQDGHQWTLTFQNGKVKVFGQNDSTFPVNSLDPLLTRYLRIHQPSWVHOIALRM 2340
Db
2281 ISSSQDGHQWTLTFQNGKVKVFGQNDSTFPVNSLDPLLTRYLRIHQPSWVHOIALRM 2340
QY
2341 EVLGCQAQDLY 2351
Db
2341 EVLGCQAQDLY 2351

RESULT 2

5171844-2

; Patent No. 5171844

; APPLICANT: VAN OOVEN, ALBERT J. J.; PANNEKOEK, HANS; VERBEET,

; MARTINUS P.; VAN LEEN, ROBERT W.

; TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY

; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS

; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM

; NUMBER OF SEQUENCES: 12

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/205,226
; FILING DATE: 10-JUN-1988
; SEQ ID NO:2
; LENGTH: 2351
; 517844-2

Query Match
Best Local Similarity 99.9%; Score 12404; DB 6; Length 2351;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOIELSTCFFLCLLRCFCSATRRYYLGAVELSWDYMOSDLGSLPVDARPPRPVPSPPFN 60
DB 1 MOIELSTCFFLCLLRCFCSATRRYYLGAVELSWDYMOSDLGSLPVDARPPRPVPSPPFN 60

QY 61 TSVVYKTLFVEFTDHLNIAKPRPPWGLGFTTQAEVDFVWVITLKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLNIAKPRPPWGLGFTTQAEVDFVWVITLKNMASHPVSLHAV 120

QY 121 GVSVMKASGAEYDDOTSOREKEDDKVFPGGSHTVVWVLKXGPMASDPLCLCTSYLSH 180
DB 121 GVSVMKASGAEYDDOTSOREKEDDKVFPGGSHTVVWVLKXGPMASDPLCLCTSYLSH 180

QY 131 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKPILLFAVPDEGKSWHSETKNSLMQDRD 240
DB 131 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKPILLFAVPDEGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPMHTVNGVYVNSPLCLTGCHEKSVVWVHVI GMGTTPEVHSIFLECHTFLVRNH 300
DB 241 AASARAWPMHTVNGVYVNSPLCLTGCHEKSVVWVHVI GMGTTPEVHSIFLECHTFLVRNH 300

QY 301 ROASLEISPIFTLTAQTLMLDGLQFLLSCHISHOHDMGEAVKYVDSQPEEPOLRMKNE 360
DB 301 ROASLEISPIFTLTAQTLMLDGLQFLLSCHISHOHDMGEAVKYVDSQPEEPOLRMKNE 360

QY 361 EADYDDDLTQSEMDVVRFPDDNSPSFTQIRSVAXKPKTWVHYIAABEEDWDYAPLVIA 420
DB 361 EADYDDDLTQSEMDVVRFPDDNSPSFTQIRSVAXKPKTWVHYIAABEEDWDYAPLVIA 420

QY 421 PDDRSYKQYLNNGPQRI GRKYKVRWAYTDETEKTRAEACHESGILGPLYGVGDTL 480
DB 421 PDDRSYKQYLNNGPQRI GRKYKVRWAYTDETEKTRAEACHESGILGPLYGVGDTL 480

QY 481 LIIFKQASRPNIYPHGITDVRPLYSRLPKGVXHLKDFPILPCEIIPKYKWTVVEDGP 540
DB 481 LIIFKQASRPNIYPHGITDVRPLYSRLPKGVXHLKDFPILPCEIIPKYKWTVVEDGP 540

QY 541 TKSDPRCLTRYSSVFNMERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDE 600
DB 541 TKSDPRCLTRYSSVFNMERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVFDE 600

QY 601 NRSWYLTENIORELPNPAQVQLEDPEFOASNTMHSGVVDLSLQSVCLHEVAYWYILS 660
DB 601 NRSWYLTENIORELPNPAQVQLEDPEFOASNTMHSGVVDLSLQSVCLHEVAYWYILS 660

QY 661 ICAQTDPLSVFPFGYTFKHKVYEDTLTFPFGSETVFMSENPGILWILGCHNSDFNRG 720
DB 661 ICAQTDPLSVFPFGYTFKHKVYEDTLTFPFGSETVFMSENPGILWILGCHNSDFNRG 720

QY 721 MTALKVSSCDKNTGDDYEDSDVILSAVLKXNAIEPRSPSONSRHPSTROKQFNATTI 780
DB 721 MTALKVSSCDKNTGDDYEDSDVILSAVLKXNAIEPRSPSONSRHPSTROKQFNATTI 780

QY 781 PENDTEKTDPMFAHRTPEKIQNVSSDDLMLLROSPTPHGLSLSDLOEAKVETFSDDPS 840
DB 781 PENDTEKTDPMFAHRTPEKIQNVSSDDLMLLROSPTPHGLSLSDLOEAKVETFSDDPS 840

QY 841 PCADSNNSLSGEMTHFRPQLHSHGDMVTFPPBSGLQRLNEKLGTTAAATELKKLPKVSST 900
DB 841 PCADSNNSLSGEMTHFRPQLHSHGDMVTFPPBSGLQRLNEKLGTTAAATELKKLPKVSST 900

QY 901 SNNLISSTIPSONLAAGTNTSSLGPPSPVHYDSQDITTLFGKKSPLTESGGPISLSEE 960
DB 901 SNNLISSTIPSONLAAGTNTSSLGPPSPVHYDSQDITTLFGKKSPLTESGGPISLSEE 960

QY 961 NDSKLLSGLMNSQESSWGKNVSTESGRUFGKRAHGPALLITKDNALFKVYSILKLTN 1020
DB 961 NDSKLLSGLMNSQESSWGKNVSTESGRUFGKRAHGPALLITKDNALFKVYSILKLTN 1020

QY 1021 KTSNNSATNRKTHIDGPSLLJTENSPSVWONLLESDETEKKTPTPLIHDRMLKDNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLJTENSPSVWONLLESDETEKKTPTPLIHDRMLKDNATLRL 1080

QY 1081 NMSNKTSSKNMFWQKKGPIPPDAQNPMSPFKMLFLPESARWQORCHGKNSLNSG 1140
DB 1081 NMSNKTSSKNMFWQKKGPIPPDAQNPMSPFKMLFLPESARWQORCHGKNSLNSG 1140

QY 1141 QGSPKQVSLGPEKSVGQNFLESEKNKVYVKGGEFTKDVGLKEMVFPSSRNLFUTNLND 1200
DB 1141 QGSPKQVSLGPEKSVGQNFLESEKNKVYVKGGEFTKDVGLKEMVFPSSRNLFUTNLND 1200

QY 1201 LHENNTNQBKIKBEEIEKKETLLOENVVLPOIHTVGTGKNFMKNLFLLSTRQNVGSGYD 1260
DB 1201 LHENNTNQBKIKBEEIEKKETLLOENVVLPOIHTVGTGKNFMKNLFLLSTRQNVGSGYD 1260

QY 1261 GAYAPVLQDFRSLNDSNTRTKKHTAHFSGKGBEENLEGLGNQTKQIVKCYACTTISNPT 1320
DB 1261 GAYAPVLQDFRSLNDSNTRTKKHTAHFSGKGBEENLEGLGNQTKQIVKCYACTTISNPT 1320

QY 1321 SQQNFVTOQRKRALQKRLPLEETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQQNFVTOQRKRALQKRLPLEETELEKRIIVDDTSTQWSNMKHLTPSTLTQIDYNEKE 1380

QY 1381 KGATQSPSLSCPLRSHSIPQANRSPPLIAKVSFPSPRIPIYLTREVLPQDNSSHLPAASY 1440
DB 1381 KGATQSPSLSCPLRSHSIPQANRSPPLIAKVSFPSPRIPIYLTREVLPQDNSSHLPAASY 1440

QY 1441 RKXDSGVQESSHFIQGAQKNNLSAILTLEMTGQOREVGSISLTSATNSVTYKVENTVLP 1500
DB 1441 RKXDSGVQESSHFIQGAQKNNLSAILTLEMTGQOREVGSISLTSATNSVTYKVENTVLP 1500

QY 1501 KPDLFKTSGKVELLPKHVYIYOKDLPPTETSNQSGHLDLVREGSLLOQTEGAIKWNEANRP 1560
DB 1501 KPDLFKTSGKVELLPKHVYIYOKDLPPTETSNQSGHLDLVREGSLLOQTEGAIKWNEANRP 1560

QY 1561 GKVPFLRVATSSAKTPSKLLDPLAWNHYGTQIPKSEWKSQESKPEKTAPEKKDTILSL 1620
DB 1561 GKVPFLRVATSSAKTPSKLLDPLAWNHYGTQIPKSEWKSQESKPEKTAPEKKDTILSL 1620

QY 1621 NACESNEAAAINEGQNKPELEVWAKQGRTERLCSQNPVVLKXHQREIITTTLQSDOEE 1680
DB 1621 NACESNEAAAINEGQNKPELEVWAKQGRTERLCSQNPVVLKXHQREIITTTLQSDOEE 1680

QY 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSPOKKTGRHYPIAAVERLWDYMGSSSPHVLNR 1740
DB 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSPOKKTGRHYPIAAVERLWDYMGSSSPHVLNR 1740

QY 1741 AQSSVPOFKKVFQEPDQSGFTQPLRGELNEHLLGPGVIRAEVDENIMVTPRNOASR 1800
DB 1741 AQSSVPOFKKVFQEPDQSGFTQPLRGELNEHLLGPGVIRAEVDENIMVTPRNOASR 1800

QY 1801 PYSYSSLIISYEDQROGAEPKRNKVFAPNETKITFYKQVQHMAPTKQEFDCWAYFSDV 1860
DB 1801 PYSYSSLIISYEDQROGAEPKRNKVFAPNETKITFYKQVQHMAPTKQEFDCWAYFSDV 1860

QY 1861 DLEKDVHSGILGPLLCHVTNLPNAGROVTVQBFALFFITFDTKSWYFENNERCRA 1920
DB 1861 DLEKDVHSGILGPLLCHVTNLPNAGROVTVQBFALFFITFDTKSWYFENNERCRA 1920

QY 1921 PCNTQMDPTFKENYRFAINGYIMDLTLPGLVMAQDORIPWYLLSMGNSNENIHHSFSGH 1980
DB 1921 PCNTQMDPTFKENYRFAINGYIMDLTLPGLVMAQDORIPWYLLSMGNSNENIHHSFSGH 1980

QY 1981 VFTYRKKEEKYKMALYNLYPGVFETVEMLPKAGIRWVECLIGEHLHAGMSTFLVYSNKC 2040
DB 1981 VFTYRKKEEKYKMALYNLYPGVFETVEMLPKAGIRWVECLIGEHLHAGMSTFLVYSNKC 2040

```

QY 2041 QTPGASGHIRDFOITASQYQWAPKIALRHYSINAWSTKPSWKVDLAPMII 2100
Db 2041 QTPGASGHIRDFOITASQYQWAPKIALRHYSINAWSTKPSWKVDLAPMII 2100
QY 2101 HGKTCQAKQKSSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGKENIEN 2160
Db 2101 HGKTCQAKQKSSLYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGKENIEN 2160
QY 2161 PPIIABYIILPHPHYGIRSLRMLGCDLNSCSMPGLGHSKAI SDAQTASSYFTNMF 2220
Db 2161 PPIIABYIILPHPHYGIRSLRMLGCDLNSCSMPGLGHSKAI SDAQTASSYFTNMF 2220
QY 2221 TWSPSKARLHLOGRSNAWRPQVNNPKEMLOQVDFQKMKVGTGVTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAWRPQVNNPKEMLOQVDFQKMKVGTGVTQGVKSLLTSMYVKEFL 2280
QY 2281 ISSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLZIHFFQSWHQAIRLM 2340
Db 2281 ISSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLLTRYLZIHFFQSWHQAIRLM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 3
US-08-366-851A-2
; Sequence 2, Application US/08366851A
; Patent No. 5681746
; GENERAL INFORMATION:
; APPLICANT: Bodner, Mordechai
; APPLICANT: De Polo, Nicolas J.
; APPLICANT: Hsu, David Chi-Tang
; APPLICANT: Chang, Steven
; TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Viagene, Inc.
; CITY: 11055 Roselle Street
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,851A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chambers, Daniel M.
; REGISTRATION NUMBER: 34,561
; REFERENCE/DOCKET NUMBER: 930049,438
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 452-1288
; TELEFAX: (619) 452-2616
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-366-851A-2
Query Match: 99.9%; Score 12401; DB 1; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQIELSTCFELCLLRCFSAIRRYLGAVELSWDYNQSDLGELPVDARPPVPKSPFPN 60

Db 1 MQIELSTCFELCLLRCFSAIRRYLGAVELSWDYNQSDLGELPVDARPPVPKSPFPN 60
QY 61 TSVYVKTLFVEFTDELFNIAKPRPPMGLGPTTQAQEVYDVTWITLKNWASHPVSLHAV 120
Db 61 TSVYVKTLFVEFTDELFNIAKPRPPMGLGPTTQAQEVYDVTWITLKNWASHPVSLHAV 120
QY 121 GYSYKASBGAIDYDQTSOREKEDDKVPFGGSHYVWVLKENGPMASDPLCLTYSYLSH 180
Db 121 GYSYKASBGAIDYDQTSOREKEDDKVPFGGSHYVWVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVXDLNSGLLGALLVCREGSLAKEKTQTLHKFTILLPAVDEGKSWHSETKNSLMQDSD 240
Db 181 VDLVXDLNSGLLGALLVCREGSLAKEKTQTLHKFTILLPAVDEGKSWHSETKNSLMQDSD 240
QY 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHURKSVYVHVIENGTTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVNGVYVNRSLPGLIGCHURKSVYVHVIENGTTPEVHSIFLEGHTFLVRNH 300
QY 301 RCASLEISPIITELTAQTLMDLGOFLSCHISSHODGMEAYKVYDSCPEEFQLEMKNE 360
Db 301 RCASLEISPIITELTAQTLMDLGOFLSCHISSHODGMEAYKVYDSCPEEFQLEMKNE 360
QY 361 EAEDYDDDLTDSMDVVRPDDNPSFQIRSVAKKHPTWVHYTAAABEEDWDYAPLVIA 420
Db 361 EAEDYDDDLTDSMDVVRPDDNPSFQIRSVAKKHPTWVHYTAAABEEDWDYAPLVIA 420
QY 421 PDDRSYKQYLNNGPQIRGRYKVRPMAYTDETFKTREAIQHESGILGLLYGEGDTL 480
Db 421 PDDRSYKQYLNNGPQIRGRYKVRPMAYTDETFKTREAIQHESGILGLLYGEGDTL 480
QY 481 LIIFKNOASRPNTYPHGITDVRPLYSRRLPGVXELKDFPLPGELPKYKWTVEQGP 540
Db 481 LIIFKNOASRPNTYPHGITDVRPLYSRRLPGVXELKDFPLPGELPKYKWTVEQGP 540
QY 541 TXSDPRCLTRYYSFVNMERDLASGLIGPLICYKESVDQGNQIMSDKRNVLFSVDE 600
Db 541 TXSDPRCLTRYYSFVNMERDLASGLIGPLICYKESVDQGNQIMSDKRNVLFSVDE 600
QY 601 NRSWYLTENIQRPFPNPAQVLEDEPFOASNTMHSINGVYVFDLSQLSVCLHVAWYILS 660
Db 601 NRSWYLTENIQRPFPNPAQVLEDEPFOASNTMHSINGVYVFDLSQLSVCLHVAWYILS 660
QY 661 IGAOTDFLSVPESGYTFXHKVYVETLTLFPESGTVFMSMENPGIWLGCNHSDFRNG 720
Db 661 IGAOTDFLSVPESGYTFXHKVYVETLTLFPESGTVFMSMENPGIWLGCNHSDFRNG 720
QY 721 MTALLKVSDDXNTGDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTI 780
Db 721 MTALLKVSDDXNTGDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFNATTI 780
QY 781 PENDIEKTPWFAHRTMPKIQNTSSSLLMLRQSPHGLSLSDLOEAKYETSDSDPS 840
Db 781 PENDIEKTPWFAHRTMPKIQNTSSSLLMLRQSPHGLSLSDLOEAKYETSDSDPS 840
QY 841 PGALDSNNLSMTHFRPQLHSHGDMVTPBSGLQRLNEKLGTTAAATELKXLDKFKVST 900
Db 841 PGALDSNNLSMTHFRPQLHSHGDMVTPBSGLQRLNEKLGTTAAATELKXLDKFKVST 900
QY 901 SNNLISTIPSONLAAGTNTSSLGPPSNPVHYDSOLDTTLFGKSSPLTESGPPSLSEE 960
Db 901 SNNLISTIPSONLAAGTNTSSLGPPSNPVHYDSOLDTTLFGKSSPLTESGPPSLSEE 960
QY 961 NNDSKLLSGLMNSOESSGKNVSTESGRFLKGRAGHPALLTKDNALFKVSTSLTKTN 1020
Db 961 NNDSKLLSGLMNSOESSGKNVSTESGRFLKGRAGHPALLTKDNALFKVSTSLTKTN 1020
QY 1021 KTSNNSATNRKTHUDGPSILLIENSVPVWNT11ESDTEFKVTPLTHDRMLMDKNATLRL 1080
Db 1021 KTSNNSATNRKTHUDGPSILLIENSVPVWNT11ESDTEFKVTPLTHDRMLMDKNATLRL 1080
QY 1081 NNSKNTTSSKNMVMVQKKEGP1PPDAQNPDMSFFKMLFLPESARW1QRTHGKNSLNSG 1140

1081 NHMSKTTSSKNMVMQCKEGP1PPDAQNPMQSFPMFLPESARW1QRTKGNLSNG 1140
1141 QGSPKQVLSLGPKEVGEONFLSEKXKVVGKGFETDVGKEMVFPSSRNLFNTLND 1200
1141 QGSPKQVLSLGPKEVGEONFLSEKXKVVGKGFETDVGKEMVFPSSRNLFNTLND 1200
1201 LHENNTNQEKKIQEIEIEKKTILQENVVLPOHTVTGTFKNFMKNLFLSTRQNVGSYD 1260
1201 LHENNTNQEKKIQEIEIEKKTILQENVVLPOHTVTGTFKNFMKNLFLSTRQNVGSYD 1260
1261 GAYAPVLDQFRSLNSTNETKHTAHFSGKGBEENLEGNGOTKQIVKACITRISPT 1320
1261 GAYAPVLDQFRSLNSTNETKHTAHFSGKGBEENLEGNGOTKQIVKACITRISPT 1320
1321 SQCNFVQSRKALKQFRPLBETELEK311VDDTSTOWSKMKHLTPSTLTQIDYNEKE 1380
1321 SQCNFVQSRKALKQFRPLBETELEK311VDDTSTOWSKMKHLTPSTLTQIDYNEKE 1380
1381 KGAIQSPUSDCILTRSHSPQANRSPPLPIAKVSSPPIRPIYLRVLFQDNSSHLPAASY 1440
1381 KGAIQSPUSDCILTRSHSPQANRSPPLPIAKVSSPPIRPIYLRVLFQDNSSHLPAASY 1440
1441 RKDQSGVQESSHFLQAKKNNLSLAILLEXTGDQREVGSLGTSATNSVYKKVNTVLP 1500
1441 RKDQSGVQESSHFLQAKKNNLSLAILLEXTGDQREVGSLGTSATNSVYKKVNTVLP 1500
1501 KPDLPKTSKVBLLPKVHIYQKDLFPETSTNSGPHLDLVEGSLQGTGGAIKWNEANRP 1560
1501 KPDLPKTSKVBLLPKVHIYQKDLFPETSTNSGPHLDLVEGSLQGTGGAIKWNEANRP 1560
1561 GKVPFLRVATSSAKTPSKLLPLAWDNHYGTQIPKEWKSGQEKPEKTAFFKKDITLSL 1620
1561 GKVPFLRVATSSAKTPSKLLPLAWDNHYGTQIPKEWKSGQEKPEKTAFFKKDITLSL 1620
1621 NACSENHAIANEGONKPEIEVWAKQSTERI CSQNPVVKRQREITRTTLOSDEE 1680
1621 NACSENHAIANEGONKPEIEVWAKQSTERICSONPPVVKRQREITRTTLOSDEE 1680
1681 IDYDTISVEMKKEDFDIYDENOSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1681 IDYDTISVEMKKEDFDIYDENOSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1741 AQSGVPOFKKVPOFTDGSQTPLYRGELNEHLGLGPYIRAEVEDNIMVTFNQASR 1800
1741 AQSGVPOFKKVPOFTDGSQTPLYRGELNEHLGLGPYIRAEVEDNIMVTFNQASR 1800
1801 PYSFYSSLIISYEDDORQGAERKNEKNTKTYFWKHMAPTKDBFDCCKAWAYFSDV 1860
1801 PYSFYSSLIISYEDDORQGAERKNEKNTKTYFWKHMAPTKDBFDCCKAWAYFSDV 1860
1861 DLEKDVHSLGLIPLVCHTNTLNPAGRQVTVQEFALPFTIPDETQSWYFTENMERNCR 1920
1861 DLEKDVHSLGLIPLVCHTNTLNPAGRQVTVQEFALPFTIPDETQSWYFTENMERNCR 1920
1921 PCNIOMEDPTKENYRFAINGYIMDTPLGLWAQDQIRWYLLSMGSGNENIHSHPGSH 1980
1921 PCNIOMEDPTKENYRFAINGYIMDTPLGLWAQDQIRWYLLSMGSGNENIHSHPGSH 1980
1981 VFTVRKKEEYKXALYNLPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKXALYNLPGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTPGLMAGSHIRDFQITASQGYQWAPKARLHYSGSINAWSTKEPFSWIKVDLAPMII 2100
2041 QTPGLMAGSHIRDFQITASQGYQWAPKARLHYSGSINAWSTKEPFSWIKVDLAPMII 2100
2101 HGIKTQGAQXESSLIYSQFIMYSLDGKKQCTYRGNSTGTLMVFFGNVDSSGKHNIFN 2160
2101 HGIKTQGAQXESSLIYSQFIMYSLDGKKQCTYRGNSTGTLMVFFGNVDSSGKHNIFN 2160
2161 PPIIARYIRLJPHTHYSIRSTIRMLMGLDNLNSCSMPLGNESKASDAQITASSYFTNFA 2220
2161 PPIIARYIRLJPHTHYSIRSTIRMLMGLDNLNSCSMPLGNESKASDAQITASSYFTNFA 2220

QY 2221 TWSPSKARLHLQGSNANRPQNNPKWLQVDFOKTMKVGTGTTQGVKSLTSMTVKBEFL 2280
Db 2221 TWSPSKARLHLQGSNANRPQNNPKWLQVDFOKTMKVGTGTTQGVKSLTSMTVKBEFL 2280
QY 2281 ISSSDQGHQWTLFFONGKVKVFOGNQDSFTPVVNSLDPLLTTRYLRHPOSVHQAIALRM 2340
Db 2281 ISSSDQGHQWTLFFONGKVKVFOGNQDSFTPVVNSLDPLLTTRYLRHPOSVHQAIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

RESULT 4

US-08-121-202-2
; Sequence 2, Application US/08121202
; Patent No. 5563045
; GENERAL INFORMATION:
; APPLICANT: Pittman, Debra
; APPLICANT: Rehenculla, Alhawaz
; APPLICANT: Wozney, Cohn M.
; APPLICANT: Kaufman, Randal J.
; TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESS: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,202
; FILING DATE: 14-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinert, M. C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1210 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-121-202-2

Query Match 99.8%; Score 12397; DB 1; Length 2351;
Best local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPFPRVKSPPFN 60
Db 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYMOSDLGELPVDARFPFPRVKSPPFN 60
QY 61 TSVVYKTKLTFVEFTDHLFNIAKRPMPWMLLGPTIQAEVYDVTVVITLKNMASHPVSLHAV 120
Db 61 TSVVYKTKLTFVEFTDHLFNIAKRPMPWMLLGPTIQAEVYDVTVVITLKNMASHPVSLHAV 120
QY 121 GVSYWKASGAEYDDQTSQREKDDKVPFGSGHTYVMQVLKENGPMASDPLCLTYSLSH 180
Db 121 GVSYWKASGAEYDDQTSQREKDDKVPFGSGHTYVMQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDLNSGLIGALLVCRBGLAKEXTQTLHKFTLLFAVFDGKSWHSETNSLMQDRD 240

181 VDLVXDJNSGLIGALLVCRBGLAKEKTOITLHKFILLFAVDFBGSXWSSTKNSLQCD3D 240
241 AASARAWPKXHTVNGVYVNSLPGJGCHRSKVYVHVIGMTTFEVSIFLEHGTTFVRNH 300
241 AASARAWPKXHTVNGVYVNSLPLGJGCHRSKVYVHVIGMTTFEVSIFLEHGTTFVRNE 300
301 ROASLEISPIITFTTAQTLMDLQGFJLSCHISSHQHDGMEAYVKVDCBEPQLRMKNE 360
301 ROASLEISPIITFTTAQTLMDLQGFJLFCCHISHQHDGMEAYVKVDCBEPQLRMKNE 360
361 EADYDDDLTDSMDVVRPDDNSPSFIQIRSVAKKHPKTWVHYIAAEEDMDYAPJVL4 420
361 EADYDDDLTDSMDVVRPDDNSPSFIQIRSVAKKHPKTWVHYIAAEEDMDYAPJVL4 420
421 PDRSRYSQVYVNSLNGPQIRGKYKVRFMAYTDETEKTRTAEIOHESGILGPLYGVEGDTL 480
421 PDRSRYSQVYVNSLNGPQIRGKYKVRFMAYTDETEKTRTAEIOHESGILGPLYGVEGDTL 480
481 LIIFKNQAGRPYNIYPEG-TDVARJXSRLP2KGVGHLKDPPIIPGSIFFKYKWTVTVEDGP 540
481 LIIFKNQAGRPYNIYPHGTTDVARJXSRLP2KGVGHLKDPPIIPGSIFFKYKWTVTVEDGP 540
541 TKSDEPCLTRYYSFVNMRDLASGLIGPLLIICYKESVDORGNOIMSDKXNVILFSVFE 600
541 TKSDEPCLTRYYSFVNMRDLASGLIGPLLIICYKESVDORGNOIMSDKXNVILFSVFE 600
601 NRSWYLTENIQRLPNPAGVQLDPEFQASNIWHSINGVYVDSLOJLVCCLHEVAYVYILS 660
601 NRSWYLTENIQRLPNPAGVQLDPEFQASNIWHSINGVYVDSLOJLVCCLHEVAYVYILS 660
661 IGAQTDFLSVFSGYTFKHQWYEDTLTLPFSGETVFMSENPGWILGCHNSDFRNG 720
661 IGAQTDFLSVFSGYTFKHQWYEDTLTLPFSGETVFMSENPGWILGCHNSDFRNG 720
721 MTALLKVSQCDKNTGYEDSVZDIJAYLLSKNAIEPRSPNSRHPSTROKQFNATTI 780
721 MTALLKVSQCDKNTGYEDSVZDIJAYLLSKNAIEPRSPNSRHPSTROKQFNATTI 780
781 PENDIEKTPWFARHTPMKPIQNVSSDDLMLRQSPTHGLSLDLOBAKYTFPSDDPS 840
781 PENDIEKTPWFARHTPMKPIQNVSSDDLMLRQSPTHGLSLDLOBAKYTFPSDDPS 840
841 PGADSNNSLSEMTHTPQLHSGDMVFTPESGLOLRNEKLGTTAATLKLDFKVSST 900
841 PGADSNNSLSEMTHTPQLHSGDMVFTPESGLOLRNEKLGTTAATLKLDFKVSST 900
901 SNNLISTIPSDNLAAGTDNTSSLGPPMPVHYDSQIDTTLFGKSSPLTESGGLSLSEE 960
901 SNNLISTIPSDNLAAGTDNTSSLGPPMPVHYDSQIDTTLFGKSSPLTESGGLSLSEE 960
961 NNDKSLLESGLMNSQESSWGKNVSSSTESGRLFKGKAHGPALLTKDNAJFKVSI5LLKTN 1020
961 NNDKSLLESGLMNSQESSWGKNVSSSTESGRLFKGKAHGPALLTKDNAJFKVSI5LLKTN 1020
1021 KTSNNSATNRKTHIDGSPILLIENSPVQWQNILESDFEFKKVPTLIHDMMLMKNAALRL 1080
1021 KTSNNSATNRKTHIDGSPILLIENSPVQWQNILESDFEFKKVPTLIHDMMLMKNAALRL 1080
1081 NHW5NKTTS5KNMEVQCKEGPIPPDAQNPDM5RPFKYLFLPESAEWIORTHGKNSLNSG 1140
1081 NHW5NKTTS5KNMEVQCKEGPIPPDAQNPDM5RPFKYLFLPESAEWIORTHGKNSLNSG 1140
1141 QGSPKQLVSLGPEK5VEGQNF5SEKNKVVGXGEFTKDVGLKEMVFP5SRNLF5TNLDN 1200
1141 QGSPKQLVSLGPEK5VEGQNF5SEKNKVVGXGEFTKDVGLKEMVFP5SRNLF5TNLDN 1200
1201 LHENNTNHEKKIQE5IEKKETLIQENVLPOIHTVTG5KQENKKNL5LL5TRQNVGSYD 1260
1201 LHENNTNHEKKIQE5IEKKETLIQENVLPOIHTVTG5KQENKKNL5LL5TRQNVGSYD 1260
1261 GAVAPVLOQDFRSLNDSTNRTKKTAFHFSKKGEEENLEGLGNQTKQIVEXYACTTRISPNT 1320

1261 GAVAPVLOQDFRSLNDSTNRTKKTAFHFSKKGEEENLEGLGNQTKQIVEXYACTTRISPNT 1320
1321 SQONFVTOQSKAALQKFRPLPIBETELEKRIIVDDTSTQWSKMKHLTPSTLLQIDYNEKE 1380
1321 SQONFVTOQSKAALQKFRPLPIBETELEKRIIVDDTSTQWSKMKHLTPSTLLQIDYNEKE 1380
1381 KGAITQSPUSDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTVLFDONSSHLPAA5Y 1440
1381 KGAITQSPUSDCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTVLFDONSSHLPAA5Y 1440
1441 RKKDSGVQSSSHFLQAGAKNNLSLAILTLEMTGDQREVGLSGTSATNSVYKKVENTVLP 1500
1441 RKKDSGVQSSSHFLQAGAKNNLSLAILTLEMTGDQREVGLSGTSATNSVYKKVENTVLP 1500
1501 KPDLPKTSQKVBLPKVHIYQKDLPP2FTSNCSPGHLDLVBGSLLOCTGEGAKWNEANRP 1560
1501 KPDLPKTSQKVBLPKVHIYQKDLPP2FTSNCSPGHLDLVBGSLLOCTGEGAKWNEANRP 1560
1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWSQBSQKPEKTAFFKKOTILSL 1620
1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWSQBSQKPEKTAFFKKOTILSL 1620
1621 NACESNHAIAANEGONKPEIEVWAKQGRTERLCSNPPVLKRHQREIRTRTLOSQDSE 1680
1621 NACESNHAIAANEGONKPEIEVWAKQGRTERLCSNPPVLKRHQREIRTRTLOSQDSE 1680
1681 IDYDDTISVEMKKEPDIYDEDENOSPRSFOKTRHYFFIAAVERLWDYGNSSSPHVLNR 1740
1681 IDYDDTISVEMKKEPDIYDEDENOSPRSFOKTRHYFFIAAVERLWDYGNSSSPHVLNR 1740
1741 AQSGSVQPKVKKVVOFFBTGDSFTQPLRGELNEHGLGPGYIRAEVEDNIMWTFRNOASR 1800
1741 AQSGSVQPKVKKVVOFFBTGDSFTQPLRGELNEHGLGPGYIRAEVEDNIMWTFRNOASR 1800
1801 PYSFVSSLI5Y2EDQOQABPRKXNVKNETKTYKWOHHWAPKYDEPDCXAWYFSDV 1860
1801 PYSFVSSLI5Y2EDQOQABPRKXNVKNETKTYKWOHHWAPKYDEPDCXAWYFSDV 1860
1861 DLEKDVHSLIGLPLVCHTNTLNPAGROVTVQEFALPFIETKSWYFETENMERNCR4 1920
1861 DLEKDVHSLIGLPLVCHTNTLNPAGROVTVQEFALPFIETKSWYFETENMERNCR4 1920
1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSNGSNENIHSIEFSGH 1980
1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSNGSNENIHSIEFSGH 1980
1981 VFTVRKKBEBKVALNLYPGVTFETVEMLP5KAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKBEBKVALNLYPGVTFETVEMLP5KAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
2041 QTPGLWASGHIRDPQITASGOYQWAPKLARLHYSGSINAWSTKPF5SWIKVDLLAPMII 2100
2041 QTPGLWASGHIRDPQITASGOYQWAPKLARLHYSGSINAWSTKPF5SWIKVDLLAPMII 2100
2101 HGKTCGARQKPF5SYISQFIIMYSJGKKWOTYRGNSTGTLMWVFGNVDS5GKIHNI5PN 2160
2101 HGKTCGARQKPF5SYISQFIIMYSJGKKWOTYRGNSTGTLMWVFGNVDS5GKIHNI5PN 2160
2161 PPIIARYIRLHPH5YISIR5TIRMEJGCDLNSC5MPLG5K5AISDAQITASSY5TNMFA 2220
2161 PPIIARYIRLHPH5YISIR5TIRMEJGCDLNSC5MPLG5K5AISDAQITASSY5TNMFA 2220
2221 TWSPSKARLHLQGR5NANRPQVNNFKEWLQVDFQKTMKVGTQVTKGKSLT5TMVYK5FL 2280
2221 TWSPSKARLHLQGR5NANRPQVNNFKEWLQVDFQKTMKVGTQVTKGKSLT5TMVYK5FL 2280
2281 ISS5QDQCHQWTLFFQNGKVKV5QGNQDSFTPVVNSLDP5LLTRYLRIHPQ5V5VHQIALRM 2340
2281 ISS5QDQCHQWTLFFQNGKVKV5QGNQDSFTPVVNSLDP5LLTRYLRIHPQ5V5VHQIALRM 2340
2341 EYJGCEAQDLY 2351
2341 EYJGCEAQDLY 2351

RESULT 5			
5422260-1			
; Patent No. 5422260			
; APPLICANT: KAUFMAN, RANDAL J.;PITTMAN, DEBRA D.;TOOLE,JOHN J.			
; TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTEINS			
; NUMBER OF SEQUENCES: 15			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/883,936			
; FILING DATE: 15-MAY-1992			
; PRIORITY DATA:			
; APPLICATION NUMBER: 279,485			
; FILING DATE: 02-DEC-1988; 09-DEC-1986			
; APPLICATION NUMBER: 939,658			
; FILING DATE: 09-DEC-1986			
; APPLICATION NUMBER: 932,767			
; FILING DATE: 18-NOV-1986			
; APPLICATION NUMBER: 868,410			
; FILING DATE: 29-MAY-1986			
; SEQ ID NO:1:			
; LENGTH: 2351			
5422260-1			
Query Match 99.8%; Score 12388; DB 6; Length 2351;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;			
QY	1	MOELSTCFFLLCLRPCEFSATRRVYLGAVELSWDMQSDLGELPVDARPPRVPKSFFPN	60
DB	1	MOELSTCFFLLCLRPCEFSATRRVYLGAVELSWDMQSDLGELPVDARPPRVPKSFFPN	60
QY	61	TSVVYKTLFVEFTDHLFNIAKRPMPMGLGPTQARVDTVITLKNMASHPVSLHAV	120
DB	61	TSVVYKTLFVEFTDHLFNIAKRPMPMGLGPTQARVDTVITLKNMASHPVSLHAV	120
QY	121	GVSVWKASEGABYDDQTSQRKEDDKVPFGSGHTYVWQVLKENGPMASDPCLTVSYLSH	180
DB	121	GVSVWKASEGABYDDQTSQRKEDDKVPFGSGHTYVWQVLKENGPMASDPCLTVSYLSH	180
QY	181	VDLYKDLNSGLIGALLVCRGSLAKEKXTQTLHKFTLLFAVEDEGKSWHSETKNSLMQDRD	240
DB	181	VDLYKDLNSGLIGALLVCRGSLAKEKXTQTLHKFTLLFAVEDEGKSWHSETKNSLMQDRD	240
QY	241	AASARAWPKMTVNGVYVNRSLPGLIGCHRKSVYVHVICMGTTPEVHSIFLRGHTFLVRNH	300
DB	241	AASARAWPKMTVNGVYVNRSLPGLIGCHRKSVYVHVICMGTTPEVHSIFLRGHTFLVRNH	300
QY	301	QASLEISPIITFLTAQTLMDLGQFLLECHISSHQHDMGEAYVVDSCPEEPQLRMKNE	360
DB	301	QASLEISPIITFLTAQTLMDLGQFLLECHISSHQHDMGEAYVVDSCPEEPQLRMKNE	360
QY	361	EADYDODDLTSEMVDVRFDDNPSPTQIRSVAKKPKTWVHYIAAEEEDWDVAPLVA	420
DB	361	EADYDODDLTSEMVDVRFDDNPSPTQIRSVAKKPKTWVHYIAAEEEDWDVAPLVA	420
QY	421	PDDRSYKQYLNGPQRIGRKYKVRVWAYTDEFTKTEAIOHESGILGPLYGEVGDTL	480
DB	421	PDDRSYKQYLNGPQRIGRKYKVRVWAYTDEFTKTEAIOHESGILGPLYGEVGDTL	480
QY	481	LIIIFKQASRYNIYPHGITDVRPLYSRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGP	540
DB	481	LIIIFKQASRYNIYPHGITDVRPLYSRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGP	540
QY	541	TKSDPRCLTRYYSFVNMERDLASGLIGPLIICYKESVDQGNQIMSKRVNVLFSVPDE	600
DB	541	TKSDPRCLTRYYSFVNMERDLASGLIGPLIICYKESVDQGNQIMSKRVNVLFSVPDE	600
QY	601	NRSWYLTENIQRFLENPAGVQLEDPEFOASNMHSINGVYVFDLSQLSVCLHEVAYWYLS	660
DB	601	NRSWYLTENIQRFLENPAGVQLEDPEFOASNMHSINGVYVFDLSQLSVCLHEVAYWYLS	660
QY	661	IGAQTDFLSVFSGYTFKHKNVYEDTLTLFPFSGETVFMSENFGMLTGLCHNSDFNRG	720

DB	661	IGAQTDFLSVFSGYTFKHKNVYEDTLTLFPFSGETVFMSENFGMLTGLCHNSDFNRG	720
QY	721	MTALLKVSSCDKNTGDYVEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQKFNATTI	780
DB	721	MTALLKVSSCDKNTGDYVEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQKFNATTI	780
QY	781	PENDIEKTDWFAHRTMPKIQNVSSDLMLLQSPPTPHGLSLSDLOEAKYETFSDDPS	840
DB	781	PENDIEKTDWFAHRTMPKIQNVSSDLMLLQSPPTPHGLSLSDLOEAKYETFSDDPS	840
QY	841	PGAIDSNNSISEMTHFRPOLHSGDMVFTPSGQLRLNEKLGTTAAETLKKLDFKVSST	900
DB	841	PGAIDSNNSISEMTHFRPOLHSGDMVFTPSGQLRLNEKLGTTAAETLKKLDFKVSST	900
QY	901	SNNLIITIPSDNLAAGTDNTSSLGPPMPVHYDSQLODTTLFGKXSSPLTSGGPLSSEE	960
DB	901	SNNLIITIPSDNLAAGTDNTSSLGPPMPVHYDSQLODTTLFGKXSSPLTSGGPLSSEE	960
QY	961	NNDSKLJESGLMNSQESSWGKNSVSTESGRLPFKGKRAHGPALLTKDNALFKVSIISLKTN	1020
DB	961	NNDSKLJESGLMNSQESSWGKNSVSTESGRLPFKGKRAHGPALLTKDNALFKVSIISLKTN	1020
QY	1021	KTSNNSATNRKTHIDGPSLLIENSVPVQNLLESDTBFKVTPLIHDMMLMDKNATLRL	1080
DB	1021	KTSNNSATNRKTHIDGPSLLIENSVPVQNLLESDTBFKVTPLIHDMMLMDKNATLRL	1080
QY	1081	NHMKNTSSKNMNVQKKKEGPIPPDAQNPDMEFFKMLFLPESARWIQRTGKNSLNSG	1140
DB	1081	NHMKNTSSKNMNVQKKKEGPIPPDAQNPDMEFFKMLFLPESARWIQRTGKNSLNSG	1140
QY	1141	QGPSPKOLVSLGPKSVGEGNFLSEKNKVVVGKGETKQVGLKEMVPPSSNLPITWLDN	1200
DB	1141	QGPSPKOLVSLGPKSVGEGNFLSEKNKVVVGKGETKQVGLKEMVPPSSNLPITWLDN	1200
QY	1201	LHENNTNHOEKKIOEEIEKKETLIIQENVLPQIHVTGTGKTFMKNLFLLSQNVGVSVD	1260
DB	1201	LHENNTNHOEKKIOEEIEKKETLIIQENVLPQIHVTGTGKTFMKNLFLLSQNVGVSVD	1260
QY	1261	GAYAPVLQDPRSLNDSINRKYKHTAHFSKKGZEENIIEGLGNQTOIIVKCYACTRISPT	1320
DB	1261	GAYAPVLQDPRSLNDSINRKYKHTAHFSKKGZEENIIEGLGNQTOIIVKCYACTRISPT	1320
QY	1321	SOQNFVTOBSKRALQKRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE	1380
DB	1321	SOQNFVTOBSKRALQKRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE	1380
QY	1381	KGATITQSPSLDCLPRSHSIPOANRSPLPIAKVVSFPIRPIYLTTRVLFDQNSSHLPAASY	1440
DB	1381	KGATITQSPSLDCLPRSHSIPOANRSPLPIAKVVSFPIRPIYLTTRVLFDQNSSHLPAASY	1440
QY	1441	RKXDSGVQESSHFLQGAKNKLSLAITLTMTGQORVSGISGTSAITNSVTKKVENTVLP	1500
DB	1441	RKXDSGVQESSHFLQGAKNKLSLAITLTMTGQORVSGISGTSAITNSVTKKVENTVLP	1500
QY	1501	KPDLPTKSGKVELLPKHVYIKOLPPTETSNKSGPCHLDLVEGSLQGTGEGAIKWEANRP	1560
DB	1501	KPDLPTKSGKVELLPKHVYIKOLPPTETSNKSGPCHLDLVEGSLQGTGEGAIKWEANRP	1560
QY	1561	GKVPFLVATESSAKTSPKLLDPLAWNHYGTQIPKEEWSQEKSPKTAFKKXDTILSL	1620
DB	1561	GKVPFLVATESSAKTSPKLLDPLAWNHYGTQIPKEEWSQEKSPKTAFKKXDTILSL	1620
QY	1621	NACESNHAIAINEGONKPEIYVWAKQGTERTLCSONPVLKXHQREIIRTTLOSQRE	1680
DB	1621	NACESNHAIAINEGONKPEIYVWAKQGTERTLCSONPVLKXHQREIIRTTLOSQRE	1680
QY	1681	IDYDDTISVEMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGNSSSPHVLRNR	1740
DB	1681	IDYDDTISVEMKKEDFDIYDEDENQSPRSQKTRHYFIAAVERLWDYGNSSSPHVLRNR	1740
QY	1741	AQSGSVQFKKVPQBEFTDGSFTQPLYRGLNEHLGLLGPYIAAEVEDNINVTFRNQASR	1800

1741 AOSGSVPQFKKVFQBFQSTQPLRYRGELNEHGLIGLGYRAEVEDNIMVTFRQASR 1800
1801 PYSFSSLSYSEDQSQGAPRKNFKNETKTYFKVQHHMPTQDBDCKAWAYFSDV 1860
1801 PYSFSSLSYSEDQSQGAPRKNFKNETKTYFKVQHHMPTQDBDCKAWAYFSDV 1860
1861 DLEKDVHSLGLIGLLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENNERNCRA 1920
1861 DLEKDVHSLGLIGLLVCHTNTLPAHGRQVTVQEFALFFTFIDETKSWYFTENNERNCRA 1920
1921 PCNIQMEDPTFKENYAPHAINGIMDTLPGLVMAQDQIRWYLLSNGSNENIHSIHFSGH 1980
1921 PCNIQMEDPTFKENYAPHAINGIMDTLPGLVMAQDQIRWYLLSNGSNENIHSIHFSGH 1980
1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGSHLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGSHLHAGMSTLFLVYSNKC 2040
2041 QTPPLGASGHIRDFOQTASQYQGWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
2041 QTPPLGASGHIRDFOQTASQYQGWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
2101 HGKIQGAKQKSSLYISOPIIMYSDGKKWQTYRONSTGTLNVFGNDVSSGIKENIEN 2160
2101 HGKIQGAKQKSSLYISOPIIMYSDGKKWQTYRONSTGTLNVFGNDVSSGIKENIEN 2160
2161 PPIIARYIRLHPHYGIRSTRMELMGCDLNSCMPLGMESKAIQITASSYFTNVEA 2220
2161 PPIIARYIRLHPHYGIRSTRMELMGCDLNSCMPLGMESKAIQITASSYFTNVEA 2220
2221 TWSPSKARLHLOGRSNARPOVNNKEMQLQVDFQKTKMTGVTTOGVKSLLTSMVYKEFL 2280
2221 TWSPSKARLHLOGRSNARPOVNNKEMQLQVDFQKTKMTGVTTOGVKSLLTSMVYKEFL 2280
2281 ISSQDGHQWTLFFQNGKVKVQGNQDSPTPVVNSLDPPLLYRILRHQSVHQAIALRM 2340
2281 ISSQDGHQWTLFFQNGKVKVQGNQDSPTPVVNSLDPPLLYRILRHQSVHQAIALRM 2340
2341 EVLGCEAQDLY 2351
2341 EVLGCEAQDLY 2351

RESULT 6

US-07-864-0045-4
Sequence 4, Application US/07864004E
Patent No. 5364771
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
APPLICANT: Runge, Marshall S.
TITLE OF INVENTION: Hybrid Human/Forcine Factor VIII
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/864,004B
FILING DATE: 07 APRIL 1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU106
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver cDNA sequence
US-07-864-004B-4
Query Match 99.0%; Score 12293; DB 1; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 20 ATRYYLGAVELSWDMYQSDLGELPVDARPPRPVPSKSPFNTSVVYKTLFVEFTDHPN 79
Db 1 ATRYYLGAVELSWDMYQSDLGELPVDARPPRPVPSKSPFNTSVVYKTLFVEFTDHPN 60
QY 80 IAKPRPPWMLLGPITQAEVYDVVITLKNMASHPVSLHAGVSYKASEGAEYDDDSQ 139
Db 61 IAKPRPPWMLLGPITQAEVYDVVITLKNMASHPVSLHAGVSYKASEGAEYDDDSQ 120
QY 140 REKEDDKVPFGGSHTYVMQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVPFGGSHTYVMQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTQTLHKZILLPAVDEKSWHSETKNSLMQDRDAASARAPMHTVNGYVR 259
Db 181 EGS LAKEKTQTLHKZILLPAVDEKSWHSETKNSLMQDRDAASARAPMHTVNGYVR 240
QY 260 SLPLGLIGCHRSVWHVITGMGTPPEVHSIFLEGHTFLVRNHEQASLEISPTIFLAQTLL 319
Db 241 SLPLGLIGCHRSVWHVITGMGTPPEVHSIFLEGHTFLVRNHEQASLEISPTIFLAQTLL 300
QY 320 MDLQGFLLSCHISSHQHDGMEAYVKVDSCEPPEPQLRMKNNEAEADYDDDLTDSMDVVRP 379
Db 301 MDLQGFLLSCHISSHQHDGMEAYVKVDSCEPPEPQLRMKNNEAEADYDDDLTDSMDVVRP 360
QY 380 DDNSPSFIQIRSVAKKHPKTVWHVIAAEEDWDYAPLAPDDDRSYKSOYLNNGPQIG 439
Db 361 DDNSPSFIQIRSVAKKHPKTVWHVIAAEEDWDYAPLAPDDDRSYKSOYLNNGPQIG 420
QY 440 RYKVKVRFWAYTDETFKTRATQHESGILGPLLYGEVGTLLIIFKNOASRPYNTYPHGI 499
Db 421 RYKVKVRFWAYTDETFKTRATQHESGILGPLLYGEVGTLLIIFKNOASRPYNTYPHGI 480
QY 500 TDVRPLYRRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFVNN 559
Db 481 TDVRPLYRRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPKSDPRCLTRYSSFVNN 540
QY 560 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENTORFLPNPAG 619
Db 541 RLASGLIGPLLI CYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENTORFLPNPAG 600
QY 620 VQLEDPEFQASINHSINGYVFDLSQLSVCLHEVAYWYLSIGACTDFLSVFFSGYTKH 679
Db 601 VQLEDPEFQASINHSINGYVFDLSQLSVCLHEVAYWYLSIGACTDFLSVFFSGYTKH 660
QY 680 KXVYEDTTLTLPFSGETVFMGMENPGLMILGCHNSDFNRGMWTLKLVSSCDKNTGYE 739
Db 661 KXVYEDTTLTLPFSGETVFMGMENPGLMILGCHNSDFNRGMWTLKLVSSCDKNTGYE 720
QY 740 DSYEDI SAYLLSKNNAIEPRFSQNSRHPSTROKOFNATTIPENDIEKTDWFAHRTMP 799
Db 721 DSYEDI SAYLLSKNNAIEPRFSQNSRHPSTROKOFNATTIPENDIEKTDWFAHRTMP 780

QY 800 KIQNVSSDILLMLRQSPHGLSLSDLOEAKYETSPDDPSCAIDSNNSLSGEMTHFRPQ 859
DB 781 KIQNVSSDILLMLRQSPHGLSLSDLOEAKYETSPDDPSCAIDSNNSLSGEMTHFRPQ 840
QY 860 LHSQDMVFTPSGLOLRNEKLGITTAATELKKLDFKVSSTNNLITSTPSNLAAGTGN 919
DB 841 LHSQDMVFTPSGLOLRNEKLGITTAATELKKLDFKVSSTNNLITSTPSNLAAGTGN 900
QY 920 TSSLGPPSPMEVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENNDKLLBESGLNNOQESSW 979
DB 901 TSSLGPPSPMEVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENNDKLLBESGLNNOQESSW 960
QY 980 GKNVSTESGRLEKGRAGHPALLTKDNALFKVSISSLLKTNKTSNNSATNRKTHIDGPSL 1039
DB 961 GKNVSTESGRLEKGRAGHPALLTKDNALFKVSISSLLKTNKTSNNSATNRKTHIDGPSL 1020
QY 1040 LIENGPSVWONTLESDTEFKKVTPLIHDRLMDKATALLAHMSNKTSSKNMVMVQOK 1099
DB 1021 LIENGPSVWONTLESDTEFKKVTPLIHDRLMDKATALLAHMSNKTSSKNMVMVQOK 1080
QY 1100 KEGPIPPAQNEDMSPFFKMLFLPEGARWIORTHGKNSLNSGOGSPKQLVSLGPEKSVBG 1159
DB 1081 KEGPIPPAQNEDMSPFFKMLFLPEGARWIORTHGKNSLNSGOGSPKQLVSLGPEKSVBG 1140
QY 1160 QNFLSEKXKVVVGKGEFTKDVGLKEMVFPSSRNLFJTNLDNLHNNTHNQEKKIQEEIEK 1219
DB 1141 QNFLSEKXKVVVGKGEFTKDVGLKEMVFPSSRNLFJTNLDNLHNNTHNQEKKIQEEIEK 1200
QY 1220 KETLLOENVVLPOIHTVTGKMFKNLFLSTRQNVGSGYDAGAYAVIODPRLSNDSTNR 1279
DB 1201 KETLLOENVVLPOIHTVTGKMFKNLFLSTRQNVGSGYDAGAYAVIODPRLSNDSTNR 1260
QY 1280 TKKTAHFSKKEEENLEGLNGNOTQIWEKYACTTRISNTSQONFVTORSK3ALKQPEL 1339
DB 1261 TKKTAHFSKKEEENLEGLNGNOTQIWEKYACTTRISNTSQONFVTORSK3ALKQPEL 1320
QY 1340 PLEETELEKRIIIVDTSTOWSNMKEHLTPSTUQTIDYNEKEKGAITQSPSLDCLTRSHI 1399
DB 1321 PLEETELEKRIIIVDTSTOWSNMKEHLTPSTUQTIDYNEKEKGAITQSPSLDCLTRSHI 1380
QY 1400 POANSPPLIAKVSFPSPRIPLIRVLFDQNSSHLPAASYEKDQSGVQESSHFLOGAKK 1459
DB 1381 POANSPPLIAKVSFPSPRIPLIRVLFDQNSSHLPAASYEKDQSGVQESSHFLOGAKK 1440
QY 1460 NNL5AILLLEMTGQOREVSGSLGTSATNSVTYKVENTVLPKPDLPKTSKGVKELLIPKVHI 1519
DB 1441 NNL5AILLLEMTGQOREVSGSLGTSATNSVTYKVENTVLPKPDLPKTSKGVKELLIPKVHI 1500
QY 1520 YOKDLFPPTETNSGSPGHLDLVBEGSLLOQTEGAIKWEANRPKVPLRVATESSAKTFSK 1579
DB 1501 YOKDLFPPTETNSGSPGHLDLVBEGSLLOQTEGAIKWEANRPKVPLRVATESSAKTFSK 1560
QY 1580 LLDPLAWNHYGTQIPKEWKSQESKPEKTAPEKXDTILSLNACSNHAIAINESQONKP 1639
DB 1561 LLDPLAWNHYGTQIPKEWKSQESKPEKTAPEKXDTILSLNACSNHAIAINESQONKP 1620
QY 1640 ELEVTAQGRTERLCSQNPVPLKREHOREITRTTLQSDQEEIDYDDTISVEMKEDFDIY 1699
DB 1621 ELEVTAQGRTERLCSQNPVPLKREHOREITRTTLQSDQEEIDYDDTISVEMKEDFDIY 1680
QY 1700 DEDENQSPRSFOKTRHYFIAAVERLWYGMSSSPHVLNRAQSGSVQPKVQVQEBTD 1759
DB 1681 DEDENQSPRSFOKTRHYFIAAVERLWYGMSSSPHVLNRAQSGSVQPKVQVQEBTD 1740
QY 1760 GSFTQPLVKGELNEHLGLLGPVIRAEVDENIMVTFRNOASRPYSFYSSLI SYEEDQROGA 1819
DB 1741 GSFTQPLVKGELNEHLGLLGPVIRAEVDENIMVTFRNOASRPYSFYSSLI SYEEDQROGA 1800
QY 1820 EPRKNVFXNETKTYFWKVQHVMVAPTKDEDFCKAWAYTSDVDLEKDVHSGLIGPLLVCHT 1879
DB 1801 EPRKNVFXNETKTYFWKVQHVMVAPTKDEDFCKAWAYTSDVDLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFTTIFDSTKSWYFTTENMERNCRAPCNQMEDPTFKENYRPHA 1939

DB 1861 NTLNPAHGRQVTVQEFALFTTIFDSTKSWYFTTENMERNCRAPCNQMEDPTFKENYRPHA 1920
QY 1940 INGYIMDTLPGIWAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTYRKKEEYKVALYNLYP 1999
DB 1921 INGYIMDTLPGIWAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTYRKKEEYKVALYNLYP 1980
QY 2000 GVFETVEMLPSKAGIRWVECLIGEHLAGMSTLFLVYSNKCOPTPLGMASGHRDFQITAS 2059
DB 1981 GVFETVEMLPSKAGIRWVECLIGEHLAGMSTLFLVYSNKCOPTPLGMASGHRDFQITAS 2040
QY 2060 GOYGOWAPKLAELHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKQFSSLIYSQ 2119
DB 2041 GOYGOWAPKLAELHYSGSINAMSTKEPFSWIKVDLLAPMIHGIKTQGARQKQFSSLIYSQ 2100
QY 2120 FIIMYSLOGKKWQTYRGNSTGTPLMVFFGNVDSGGIKHINFPPIIARYIRLHPHTYSIRS 2179
DB 2101 FIIMYSLOGKKWQTYRGNSTGTPLMVFFGNVDSGGIKHINFPPIIARYIRLHPHTYSIRS 2160
QY 2180 TLRMELMGCDLNSCSPMLGMSKALSDAQITASSYFTNMFAWTWSPSKARLHLOQGSNAWR 2239
DB 2161 TLRMELMGCDLNSCSPMLGMSKALSDAQITASSYFTNMFAWTWSPSKARLHLOQGSNAWR 2220
QY 2240 PQVNNPKWLQVDFOKTMKVGTQVTKVKSLLTSMYKVEFLISSQDGHQWTLFPQNGKV 2299
DB 2221 PQVNNPKWLQVDFOKTMKVGTQVTKVKSLLTSMYKVEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDPPLLTLYRLRHPOSVWHQIALRMEVLGCEAQDLY 2351
DB 2281 KVFQGNQDSFTPVVNSLDPPLLTLYRLRHPOSVWHQIALRMEVLGCEAQDLY 2332

RESULT 7

US-08-251-937A-4
; Sequence 4, Application US/09251937A
; Patent No. 5583209
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marichall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/251,937A
; FILING DATE: 31-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6367
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: protein		Query Match		99.0%; Score 12293; DB 1; Length 2332;	
HYPOTHETICAL: NO		Best Local Similarity		99.9%; Pred. No. 0;	
ANTI-SENSE: NO		Matches 2329; Conservative		1; Mismatches 2; Indels 0; Gaps 0;	
FRAGMENT TYPE: N-terminal					
ORIGINAL SOURCE:					
ORGANISM: Homo sapien					
TISSUE TYPE: Liver cDNA sequence					
US-08-251-937A-4					
QY	20	ATRRYVIGAVELSWYQMSQDGLPVDARPPVPKSPFNTSVVYKTLFVDFDHLN	79		
DB	1	ATERYVIGAVELSWDQMSQDGLPVDARPPVPKSPFNTSVVYKTLFVDFTHLNF	60		
QY	80	IAXPRPMMGLGPTIOAEVYDVTWTLKXASHPVSLHAGVSYWKASEGABYDDQTSQ	139		
DB	61	IAXPRPMMGLGPTIOAEVYDVTWTLKXASHPVSLHAGVSYWKASEGABYDDQTSQ	120		
QY	140	RESEDDKVPFGSGHTYVQVLKNGPMASDPLCLITYSLSHVDLYKDLNSGLIGALLVCR	199		
DB	121	REKEDDKVPFGSGHTYVQVLKNGPMASDPLCLITYSLSHVDLYKDLNSGLIGALLVCR	180		
QY	200	EGSLAKEKQTQLHKFILLPAVFDGKSWHSETKNSLMQDQDAASARAWPKMHTVNGYVNR	259		
DB	181	EGSLAKEKQTQLHKFILLPAVFDGKSWHSETKNSLMQDQDAASARAWPKMHTVNGYVNR	240		
QY	260	SPLGLICGHRKSYVHVIGMGTTPVHSHIFLEGHTFLVRNHRQASLEISPIETFTAQTL	319		
DB	241	SPLGLICGHRKSYVHVIGMGTTPVHSHIFLEGHTFLVRNHRQASLEISPIETFTAQTL	300		
QY	320	MDLGOQLLGHISGHQDGNVAVKVDSCPEEPQLEKMNNEAEDYDDDLTDSMDVVRP	379		
DB	301	MDLGOQLLGHISGHQDGNVAVKVDSCPEEPQLEKMNNEAEDYDDDLTDSMDVVRP	360		
QY	380	DDNSPSFIQIRSVAKXGKPTWYHYIAAEEEDWDYAPLAPDDRYSKYQLANGPQRIG	439		
DB	361	DDNSPSFIQIRSVAKXGKPTWYHYIAAEEEDWDYAPLAPDDRYSKYQLANGPQRIG	420		
QY	440	RYKKVRPMAYTETPKIRSAIQHESGILGPLYGEVGDITLLIFKNQASRPNIYPHGI	499		
DB	421	RYKKVRPMAYTETPKIRSAIQHESGILGPLYGEVGDITLLIFKNQASRPNIYPHGI	480		
QY	500	TDVRPLYSRRLPGVKHLKDFPILPGSIPKYKWTIVEDGPTKSDPCLTRYYSFVNME	559		
DB	481	TDVRPLYSRRLPGVKHLKDFPILPGSIPKYKWTIVEDGPTKSDPCLTRYYSFVNME	540		
QY	560	RLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVPDENRSWLTENIQRFLENPAG	619		
DB	541	RLASGLIGPLLI CYKESVDQRGNQIMSDKRNVLFSVPDENRSWLTENIQRFLENPAG	600		
QY	620	VQLEDPEFOASNMHSNGVVFDSLOLSVCLHEVAYWYLLISGAQTDFLSVFSGYTFKH	679		
DB	601	VQLEDPEFOASNMHSNGVVFDSLOLSVCLHEVAYWYLLISGAQTDFLSVFSGYTFKH	660		
QY	680	KMYVEDTLTLFPFSGETVFMSEMPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDIYE	739		
DB	661	KMYVEDTLTLFPFSGETVFMSEMPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDIYE	720		
QY	740	DSYEDTSAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFVFAHRTMP	799		
DB	721	DSYEDTSAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFVFAHRTMP	780		
QY	800	KIQNVSSDDLMLLRQSPTHGHLSLDQAKYETTFSDPSGAISSNLSSEMTFRPQ	859		
DB	781	KIQNVSSDDLMLLRQSPTHGHLSLDQAKYETTFSDPSGAISSNLSSEMTFRPQ	840		
QY	860	LHHSQDMVFTPESGQLRLNEKLTGTTAATLKKLDFKVSSTSNLLSTIPSDNLAAGTDN	919		
DB	841	LHHSQDMVFTPESGQLRLNEKLTGTTAATLKKLDFKVSSTSNLLSTIPSDNLAAGTDN	900		
QY	920	TSSLGPPMPVHYDSQDITLLFGKXSSPLTSGGPLSLSEENNDSKLLSGLMNSQESSW	979		
DB	901	TSSLGPPMPVHYDSQDITLLFGKXSSPLTSGGPLSLSEENNDSKLLSGLMNSQESSW	960		
QY	980	GKXVSTSGRLFKGKRAHGPALLTKDNALFKVSI-SLLKTNKTSNNSATNRKTHIDGPSL	1039		
DB	961	GKXVSTSGRLFKGKRAHGPALLTKDNALFKVSI-SLLKTNKTSNNSATNRKTHIDGPSL	1020		
QY	1040	LTENSPSVWONTLESDBTFKVTPLIHDRLMDKNTALRLNEMSNKTTSSKXMEVQOK	1099		
DB	1021	LTENSPSVWONTLESDBTFKVTPLIHDRLMDKNTALRLNEMSNKTTSSKXMEVQOK	1080		
QY	1100	XEGPDPDAQNMDGFFPKMLFLPESARWIORTHGKNSLNSGQCPSPKQIVSLGPKSVESG	1159		
DB	1081	XEGPDPDAQNMDGFFPKMLFLPESARWIORTHGKNSLNSGQCPSPKQIVSLGPKSVESG	1140		
QY	1160	QNFLESKNKVVYKGEFTKDVGLKEMVFPSSNRLFTNLNLDNHNHNTNQBKKEIEIEBK	1219		
DB	1141	QNFLESKNKVVYKGEFTKDVGLKEMVFPSSNRLFTNLNLDNHNHNTNQBKKEIEIEBK	1200		
QY	1220	KETLIQENNVVLPQCHTGTGKMFKNLFLSTFRQNVGSGYDGAAPVLDQFSLNDSTNR	1279		
DB	1201	KETLIQENNVVLPQCHTGTGKMFKNLFLSTFRQNVGSGYDGAAPVLDQFSLNDSTNR	1260		
QY	1280	TKKHTAHFSKGEENLEGLGNQTKQIVKXACTTRISPTSQNFVQTSRKAALQKPL	1339		
DB	1261	TKKHTAHFSKGEENLEGLGNQTKQIVKXACTTRISPTSQNFVQTSRKAALQKPL	1320		
QY	1340	PLEETELEKRIIIVDDTSTQWSXMKHLTPSTITQIDYNEKEKGAITQSPSLCLTRSHI	1399		
DB	1321	PLEETELEKRIIIVDDTSTQWSXMKHLTPSTITQIDYNEKEKGAITQSPSLCLTRSHI	1380		
QY	1400	POANSPLEIAKVSSPSPRIPIYLVLPQDNSSHLPAASVYKKKDSGVQSSSHFLOGAKK	1459		
DB	1381	POANSPLEIAKVSSPSPRIPIYLVLPQDNSSHLPAASVYKKKDSGVQSSSHFLOGAKK	1440		
QY	1460	NNLSLAILEMTGQREVSGLSGTSATNSVTYKVENTVLPKPDLPKTSKVELLPKVHI	1519		
DB	1441	NNLSLAILEMTGQREVSGLSGTSATNSVTYKVENTVLPKPDLPKTSKVELLPKVHI	1500		
QY	1520	YOKJLPPTTSNGSPCHLDLVEGSLLOGTGEGAIKWEANRPCKVPLRLVATSSAKTPSK	1579		
DB	1501	YOKJLPPTTSNGSPCHLDLVEGSLLOGTGEGAIKWEANRPCKVPLRLVATSSAKTPSK	1560		
QY	1580	LDDPLAWDNHYGTQIPKBBWKSOEKSPEKTA PKKDTILSLNACSNHAI AAINBQONKP	1639		
DB	1561	LDDPLAWDNHYGTQIPKBBWKSOEKSPEKTA PKKDTILSLNACSNHAI AAINBQONKP	1620		
QY	1640	EIEVTWAKQGRTERLCSQNPVVKRQREITRTTILQSDQEEIDYDDTISVENMKKEDFDY	1699		
DB	1621	EIEVTWAKQGRTERLCSQNPVVKRQREITRTTILQSDQEEIDYDDTISVENMKKEDFDY	1680		
QY	1700	DDENOSPRSFOKTHYFI AAVERLWDYGMSSSHVLNRACSGSVPOPKVVFQEFDD	1759		
DB	1681	DDENOSPRSFOKTHYFI AAVERLWDYGMSSSHVLNRACSGSVPOPKVVFQEFDD	1740		
QY	1760	GSFTQPLYRGELNEHLGLGPIYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDORQGA	1819		
DB	1741	GSFTQPLYRGELNEHLGLGPIYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDORQGA	1800		
QY	1820	EPKKNPVKNENKTYFWKVOHHMAPTKDBFDCAKAYFSDVLEKDVHSGSLGPLLVCHT	1879		
DB	1801	EPKKNPVKNENKTYFWKVOHHMAPTKDBFDCAKAYFSDVLEKDVHSGSLGPLLVCHT	1860		
QY	1880	NTLNPAHGQVTVQSFALPFTTIFDETCKSWYFTENMERNCRAPCNQMEDPTFKNYRPHA	1939		
DB	1861	NTLNPAHGQVTVQSFALPFTTIFDETCKSWYFTENMERNCRAPCNQMEDPTFKNYRPHA	1920		
QY	1940	INGYIMDTLPLGLVMAQDQIRIWRVLLSMGNSNENHSHIFSGHVFTVRKKEEYKWAYLYNP	1999		
DB	1921	INGYIMDTLPLGLVMAQDQIRIWRVLLSMGNSNENHSHIFSGHVFTVRKKEEYKWAYLYNP	1980		
QY	2000	GVFETVEMLPKSKAGIWRVBECLIGEHLHAGMSTLFLVYSNKCQTPPLGMASGHIRDPQITAS	2059		

1981 CVFTEVEMLPKAGIWRVECLGELHAGMSTLFLVYSNKCQOTPLGMASGHIROFQITAS 2040
2060 QOYGOWAPKLAARLHYSGGINASTKEPESWIKVDLLAPMIHIGIKTOGAROKFSSLYISQ 2119
2041 QOYGOWAPKLAARLHYSGGINASTKEPESWIKVDLLAPMIHIGIKTOGAROKFSSLYISQ 2100
2120 FIIMYSLGKKWQOTYRGNSTGLMVFPGNVDSGGIKHFNPPIIARIYIRLHPHYSIRS 2179
2101 FIIMYSLGKKWQOTYRGNSTGLMVFPGNVDSGGIKHFNPPIIARIYIRLHPHYSIRS 2160
2180 TLRMELMGCDLNSCSWPLMGSKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAR 2239
2161 TLRMELMGCDLNSCSWPLMGSKAISDAQITASSYFTNNFATWSPSKARLHLQGRSNAR 2220
2240 POWNPKEWLOVDPOKTKWKTGVITTOGVKSLTSMYKKEFLISSQDGHOWTLFPQNGKV 2299
2221 POWNPKEWLOVDPOKTKWKTGVITTOGVKSLTSMYKKEFLISSQDGHOWTLFPQNGKV 2280
2300 KVFQGNQDSFTPVNSLDDPPLITRYLRIHPQSWHQAIALRMEVLGCEAQDLY 2351
2281 KVFQGNQDSFTPVNSLDDPPLITRYLRIHPQSWHQAIALRMEVLGCEAQDLY 2332

RESULT 8
US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Bollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Babst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence

Query Match
Best Local Similarity 99.0%; Score 12293; DB 1; Length 2332;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 20 ATRYYLGAVELSDYQSDLGELPVDARFPFRVPKSPFNTSVVYKTLFVEETDHLFN 79
DB 1 ATRYYLGAVELSDYQSDLGELPVDARFPFRVPKSPFNTSVVYKTLFVEETDHLFN 60
QY 80 IAKRPPWMLLGGTIOAEVYDVTWITLKNMASHPVSLHVGVSYWKASGEAEVDDQTSQ 139
DB 61 IAKRPPWMLLGGTIOAEVYDVTWITLKNMASHPVSLHVGVSYWKASGEAEVDDQTSQ 120
QY 140 REKDDKVPFGSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKDDKVPFGSHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKTQTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKGHTVNGYVR 259
DB 181 EGSIAKEKTQTLHKFILLFAVDFDEGKSWHSETKNSLMQDRDAASARAWPKGHTVNGYVR 240
QY 260 SLPLIGCHRSKSVYWHVIGMGTTEPEVHSIFLEGHTPLVRNHRQASLEISPTIFLTAQTL 319
DB 241 SLPLIGCHRSKSVYWHVIGMGTTEPEVHSIFLEGHTPLVRNHRQASLEISPTIFLTAQTL 300
QY 320 MDLGOFLLSCHISSHOHDGMEAYVKVDSCPEEPOLRMKNNEEAEDYDDLTDSEMDVVRP 379
DB 301 MDLGOFLLSCHISSHOHDGMEAYVKVDSCPEEPOLRMKNNEEAEDYDDLTDSEMDVVRP 360
QY 380 DDNSPSFQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRYSYKSYLNNGPQIRG 439
DB 361 DDNSPSFQIRSVAKKHPKTWVHYIAAEEDWDYAPLVLAPDDRYSYKSYLNNGPQIRG 420
QY 440 RYKVKRFMAYTDETFKTBRAIOHESGILGPLLGEVGDITLLIFKQASPYMYPHGI 499
DB 421 RYKVKRFMAYTDETFKTBRAIOHESGILGPLLGEVGDITLLIFKQASPYMYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKDPFLLPGBIFKWTVTVEDGPTKSDPPRCLITYYSFVNME 559
DB 481 TDVRPLYSRRLPKGVKHLKDPFLLPGBIFKWTVTVEDGPTKSDPPRCLITYYSFVNME 540
QY 560 RDLASGLIGPLLI CYKESVDQGRNQIMSDKRNVLFSVDFDNRSWYLTENTQRFELPDAG 619
DB 541 RDLASGLIGPLLI CYKESVDQGRNQIMSDKRNVLFSVDFDNRSWYLTENTQRFELPDAG 600
QY 620 VOLEDPEFQASNIHNSINGYVFDLSQLSVCLHEVAYWILLISGATDPLSFVSFSGYTFKH 679
DB 601 VOLEDPEFQASNIHNSINGYVFDLSQLSVCLHEVAYWILLISGATDPLSFVSFSGYTFKH 660
QY 680 KMVYEDTTLTLPFSGETVFMGMENPGLWILGCHNSDFNRNGMTALLKVSSCDKNTGDYFE 739
DB 661 KMVYEDTTLTLPFSGETVFMGMENPGLWILGCHNSDFNRNGMTALLKVSSCDKNTGDYFE 720
QY 740 DSYEDISAYLLSKNAIEPRSFQNSRHPSTROQFNATITPENDIEKTDWFAHRTMP 799
DB 721 DSYEDISAYLLSKNAIEPRSFQNSRHPSTROQFNATITPENDIEKTDWFAHRTMP 780
QY 800 KIQNVSSSDLLMLLRQSPHGLSISDLOEAKYETFDSDPGALDSNNLSSEMTFRPQ 859
DB 781 KIQNVSSSDLLMLLRQSPHGLSISDLOEAKYETFDSDPGALDSNNLSSEMTFRPQ 840
QY 860 LHSQDMVFTPESGIQLRLNEKLGTTAATELKLDFKVSSTSNLITIPSDNLAAGTDN 919
DB 841 LHSQDMVFTPESGIQLRLNEKLGTTAATELKLDFKVSSTSNLITIPSDNLAAGTDN 900
QY 920 TSSLGPPSPMPVHYDSQDITTLFGKSSPLTSGGPLSLSENNDSKLLIESGLMNSQBSW 979
DB 901 TSSLGPPSPMPVHYDSQDITTLFGKSSPLTSGGPLSLSENNDSKLLIESGLMNSQBSW 960
QY 980 GKNVSTESGELPKGKAHGPALLTKONALPKVSIISLTKNTKTNNSATNKRTHIDGSL 1039
DB 961 GKNVSTESGELPKGKAHGPALLTKONALPKVSIISLTKNTKTNNSATNKRTHIDGSL 1020

1040 LIENSPVWQWILESDTEFKVTPLIHDMMLDKNATLRLNHNKNTTSSKNMENVQOK 1099
1021 LIENSPVWQWILESDTEFKVTPLIHDMMLDKNATLRLNHNKNTTSSKNMENVQOK 106C
1100 KEGPIPPDAQNDPMFFKMLFLPESARWIQTHGKNSLNSGGPSPKQVLSLGPSPKSVBG 1159
1081 KEGPIPPDAQNDPMFFKMLFLPESARWIQTHGKNSLNSGGPSPKQVLSLGPSPKSVBG 1140
1160 QNFLSEKQKVVVGKEFTKDVGLKMWPPSSNLFITNLDNHNHNTNHEKKEKKEE-EK 1219
1141 QNFLSEKQKVVVGKEFTKDVGLKMWPPSSNLFITNLDNHNHNTNHEKKEKKEE-EK 1200
1220 KETLIQENNVLPQIHTVTGTNFMKMLFLLSTRQNVESYDGAAPVLQDFSLNDSNTR 1279
1201 KETLIQENNVLPQIHTVTGTNFMKMLFLLSTRQNVESYDGAAPVLQDFSLNDSNTR 1260
1280 TKKHTAHSKKGEBENLEGLGNOTKQIVKXACTRISNTSOQNFVTOSSKALQKPL 1339
1261 TKKHTAHSKKGEBENLEGLGNOTKQIVKXACTRISNTSOQNFVTOSSKALQKPL 1320
1340 PLEETELEKRIIVDDTSTOWSNKMHLPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
1321 PLEETELEKRIIVDDTSTOWSNKMHLPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
1400 POANRSPPIAKVSSPPIRPIYLRVLQDNSSHLPAASYPKDSGVOESSHFLQGAKK 1459
1381 POANRSPPIAKVSSPPIRPIYLRVLQDNSSHLPAASYPKDSGVOESSHFLQGAKK 1440
1460 NNLISAILTLEMTGPOREVGSLGTATNSVTYKVENVLPRKPDLPKTSKGVKVELLPKVHI 1519
1441 NNLISAILTLEMTGPOREVGSLGTATNSVTYKVENVLPRKPDLPKTSKGVKVELLPKVHI 1500
1520 YOKDLFPPTSTSGRHLDLVBSGLLQGTGEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579
1501 YOKDLFPPTSTSGRHLDLVBSGLLQGTGEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
1580 LLDPLAWDNHYGTQIPKESKWKQESPEKTAFKKDTILSNACSNHAIAINNGQKNP 1639
1561 LLDPLAWDNHYGTQIPKESKWKQESPEKTAFKKDTILSNACSNHAIAINNGQKNP 1620
1640 ELEVWAKOGRTERICSONPPVLKHOREITRTILOSDOEEIDYDDTISVENMKKEDFDIY 1699
1621 ELEVWAKOGRTERICSONPPVLKHOREITRTILOSDOEEIDYDDTISVENMKKEDFDIY 1680
1700 DEDENQSPASFOKTRHYPIAAVERLWDYGMSSSPFLNRNQAQSGSVPOFKKVVQEPFD 1759
1681 DEDENQSPASFOKTRHYPIAAVERLWDYGMSSSPFLNRNQAQSGSVPOFKKVVQEPFD 1740
1760 GSFTOPLVGEINEHLGLLGPYIRAEVDNIMVTFRNQASRPYSFYSSLSIYEEDQORQA 1819
1741 GSFTOPLVGEINEHLGLLGPYIRAEVDNIMVTFRNQASRPYSFYSSLSIYEEDQORQA 1800
1820 EPRKPFVKNETKTYFWKQHHWAPTKDFCDKAWAYFSDVLEKDVHSGLIGPLLVCHT 1879
1801 EPRKPFVKNETKTYFWKQHHWAPTKDFCDKAWAYFSDVLEKDVHSGLIGPLLVCHT 1860
1880 NTLNPAHQVTVQVQFALFTTIDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRPHA 1939
1861 NTLNPAHQVTVQVQFALFTTIDETKSWYFTENMERNCRAPCNIQMEDPTPKENYRPHA 1920
1940 INGYIMDTLPGVMAODORIRWVLLSMGNSNENIHSIHRSHVFTVTKKEEYKVALYNLYP 1999
1921 INGYIMDTLPGVMAODORIRWVLLSMGNSNENIHSIHRSHVFTVTKKEEYKVALYNLYP 1980
2000 GVFETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMAASHIRDFOITAS 2059
1981 GVFETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCOTPLGMAASHIRDFOITAS 2040
2060 GQYGWAPKARLHYSGSINAWSTKPPSWIKVLLAPMIHGIKTQGARQRFSSLYISQ 2119
2041 GQYGWAPKARLHYSGSINAWSTKPPSWIKVLLAPMIHGIKTQGARQRFSSLYISQ 2100
2120 FIIVYSIDGKKWQYRGNSGTGLMVFFGCVDSGSIKENIFNPPIIARYIRLHPHYSIRS 2179

2101 FIIVYSIDGKKWQYRGNSGTGLMVFFGCVDSGSIKENIFNPPIIARYIRLHPHYSIRS 2160
2180 TRLMELMCDLNSCSYPLGMESKAIISDAQITASSYFTNMPATWSPSKARLHLQGRSNMWR 2239
2161 TRLMELMCDLNSCSYPLGMESKAIISDAQITASSYFTNMPATWSPSKARLHLQGRSNMWR 2220
2240 PQVNNPKEWLQVDFQKTMKVTGVTITQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
2221 PQVNNPKEWLQVDFQKTMKVTGVTITQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
2300 KVFQGNQSPFTVAVNSLDPPLLTLYRLRHPSQVWHQJALRMEVLGCEAQDLY 2351
2281 KVFQGNQSPFTVAVNSLDPPLLTLYRLRHPSQVWHQJALRMEVLGCEAQDLY 2332

RESULT 9
US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 574446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; US-08-474-503-2

Query Match 99.0%; Score 12293; DB 1; Length 2332;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

20 ATERYVLGAVELSDYMDQSLGELPVDARFPPRPVPSPPNTSVYKTLFVEFTVHLFN 79
1 ATERYVLGAVELSDYMDQSLGELPVDARFPPRPVPSPPNTSVYKTLFVEFTVHLFN 60
80 IAKPRPPWMLGPGTIOAEVDTVITLKNVASHVPSHAVGVSYWKASEGAEYDDQTSQ 139
61 IAKPRPPWMLGPGTIOAEVDTVITLKNVASHVPSHAVGVSYWKASEGAEYDDQTSQ 120

QY 140 REKEDDKVPPGSGSHYVWQVILKENGPMASDPCLITYSYLSHVDLVKDLNSGLIGALLVCR 199
 DB 121 REKEDDKVPPGSGSHYVWQVILKENGPMASDPCLITYSYLSHVDLVKDLNSGLIGALLVCR 180
 QY 200 EGSIAKEKTQTLHKFILLPAVDEGKSWHSETKNSLMQDRDAASARAWPMHITVNGYVR 259
 DB 181 EGSIAKEKTQTLHKFILLPAVDEGKSWHSETKNSLMQDRDAASARAWPMHITVNGYVR 240
 QY 260 SLPGILGCHRKSVYHVGWGTTPVHSIFLGHGTFPLVRNHRQASLEISPIFLTAQTL 319
 DB 241 SLPGILGCHRKSVYHVGWGTTPVHSIFLGHGTFPLVRNHRQASLEISPIFLTAQTL 300
 QY 320 XDLGQPLFLCHISSHQHGWAEYKVDSCPEBPQLRMKNKEEAEDYDDLTDISEMDVVR 379
 DB 301 MDLQGPFLFLCHISSHQHGWAEYKVDSCPEBPQLRMKNKEEAEDYDDLTDISEMDVVR 360
 QY 380 DDNDSPTQIERSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDDRSYKQYLNNGPQRIG 439
 DB 361 DDNDSPTQIERSVAKKHPTWVHYIAAEEEDWDYAPLVLPADDDRSYKQYLNNGPQRIG 420
 QY 440 RYKXKVRFMAYUDEPFTKREAIQHESSGILGPLLYGEVGDTLIIIPKNOASRPYNIYBHI 499
 DB 421 RYKXKVRFMAYUDEPFTKREAIQHESSGILGPLLYGEVGDTLIIIPKNOASRPYNIYBHI 480
 QY 500 TDVRLYGRRLPKGVKHLKDFPILPGEIYKYKWTVTVEDGPTKSDPRLCLTRYSSFVNME 559
 DB 481 TDVRLYGRRLPKGVKHLKDFPILPGEIYKYKWTVTVEDGPTKSDPRLCLTRYSSFVNME 540
 QY 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKRVNILLFSVFEDNRSWYLHENIORFLENPAG 619
 DB 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKRVNILLFSVFEDNRSWYLHENIORFLENPAG 600
 QY 620 VQLEDEPEQASNMHSINGYVDSLSQLSVCLHEVAYWILSICAOTDFLSVFFSGYTPKH 679
 DB 601 VQLEDEPEQASNMHSINGYVDSLSQLSVCLHEVAYWILSICAOTDFLSVFFSGYTPKH 660
 QY 680 KMYEDTTLTLPFSGETVMSMENPCLMILGCHNSDFRNRGMTALLKVSSCDKNTGDYIE 739
 DB 661 KMYEDTTLTLPFSGETVMSMENPCLMILGCHNSDFRNRGMTALLKVSSCDKNTGDYIE 720
 QY 740 DSYEDISAYLLSKMNAIPERSQNSRHPSTQKQFNATTIPENDIEKTDPMFAHRTMP 799
 DB 721 DSYEDISAYLLSKMNAIPERSQNSRHPSTQKQFNATTIPENDIEKTDPMFAHRTMP 780
 QY 800 KIQNVSSDLMMLRQSPFPHGLSLSDLOEAKYETFSDDPSGALDSNNLSSEMTFPRQ 859
 DB 781 KIQNVSSDLMMLRQSPFPHGLSLSDLOEAKYETFSDDPSGALDSNNLSSEMTFPRQ 840
 QY 860 LHHSGDMVFTPESGLQRLNEKLGTTAATELKKLDFKVSSTSNLIISTIPSDNLAAGTDN 919
 DB 841 LHHSGDMVFTPESGLQRLNEKLGTTAATELKKLDFKVSSTSNLIISTIPSDNLAAGTDN 900
 QY 920 TSSLGPPNPNVHYDSQDLDTTLFGKXSSPLTESGGPLSISENNDSKLLSEGLMNSQESSW 979
 DB 901 TSSLGPPNPNVHYDSQDLDTTLFGKXSSPLTESGGPLSISENNDSKLLSEGLMNSQESSW 960
 QY 980 GKNVSSSTSGRLFKGKRAHGPAALLTKDNALFKVSI SLLKTKNTKNSNATNRKTHIDGPSL 1039
 DB 961 GKNVSSSTSGRLFKGKRAHGPAALLTKDNALFKVSI SLLKTKNTKNSNATNRKTHIDGPSL 1020
 QY 1040 LIENSPTSQWONILEDTEPKKTPPLIHDRMLMDKXATALRLNHNKNTTSKQKMEVQOK 1099
 DB 1021 LIENSPTSQWONILEDTEPKKTPPLIHDRMLMDKXATALRLNHNKNTTSKQKMEVQOK 1080
 QY 1100 KEGP-PPDAQNDPMSFFKMLRPLPESARWIORTHGKNSLNSGGPSPKQIVSLGPESKVEG 1159
 DB 1081 KEGP-PPDAQNDPMSFFKMLRPLPESARWIORTHGKNSLNSGGPSPKQIVSLGPESKVEG 1140
 QY 1160 QNFLEKKNVVVGKEFTKDVGLKEMVFPSSRNLEFTNLNLHNNTNENQEKKIOBIEK 1219
 DB 1141 QNFLEKKNVVVGKEFTKDVGLKEMVFPSSRNLEFTNLNLHNNTNENQEKKIOBIEK 1200

QY 1220 KETLIQENNVLPQJHTVTGKNFMKNLFLSTRONVEGSDYGAAPVLODFRSLNDSTNR 1279
 DB 1201 KETLIQENNVLPQJHTVTGKNFMKNLFLSTRONVEGSDYGAAPVLODFRSLNDSTNR 1260
 QY 1280 TKKHTAHFSKXGEENLEGLNQTKQIIVEKYACTTRISPTNSQQNFVTRQSKRALQFRL 1339
 DB 1261 TKKHTAHFSKXGEENLEGLNQTKQIIVEKYACTTRISPTNSQQNFVTRQSKRALQFRL 1320
 QY 1340 PLEBTELEKRIIVDDTSTQWSKNMKHLTSTLTQTDYNEKEKGAITQSPSLDCLTRSHSI 1399
 DB 1321 PLEBTELEKRIIVDDTSTQWSKNMKHLTSTLTQTDYNEKEKGAITQSPSLDCLTRSHSI 1380
 QY 1400 POANESPIPIAKVSSFPPIRPIYILTRVLFODNSSHLPAASVRKDKSGVQESSHFLQGAKK 1459
 DB 1381 POANESPIPIAKVSSFPPIRPIYILTRVLFODNSSHLPAASVRKDKSGVQESSHFLQGAKK 1440
 QY 1460 NNLSLAILTLEMTGQREVGSIGTSATNSVYKVKVENTVLPKPDLPKTSKGYELLPKVHI 1519
 DB 1441 NNLSLAILTLEMTGQREVGSIGTSATNSVYKVKVENTVLPKPDLPKTSKGYELLPKVHI 1500
 QY 1520 YOKOLPPTETSNGPSHLDLVEGSLIQTEGAIKWNBANRPKVPLVATBESSAKTFSK 1579
 DB 1501 YOKOLPPTETSNGPSHLDLVEGSLIQTEGAIKWNBANRPKVPLVATBESSAKTFSK 1560
 QY 1580 LLDPLAWDNHYGTQIIPKEEWSQESPEKTAPKKDITILSNACSNHAI AAINEGQNK 1639
 DB 1561 LLDPLAWDNHYGTQIIPKEEWSQESPEKTAPKKDITILSNACSNHAI AAINEGQNK 1620
 QY 1640 ELEVTAQGRTEELCSQNPVVLKHQBIEITRTTLOSQEBIDYDDTISVENMKEDDFIY 1699
 DB 1621 ELEVTAQGRTEELCSQNPVVLKHQBIEITRTTLOSQEBIDYDDTISVENMKEDDFIY 1680
 QY 1700 DEDENQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVPOFKKVVFOETD 1759
 DB 1681 DEDENQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVPOFKKVVFOETD 1740
 QY 1760 GSFTOPLYRGELNEHIGLGPYIRAEVEDNTMVTFRNOASRPYFSYSLISYEEDORQGA 1819
 DB 1741 GSFTOPLYRGELNEHIGLGPYIRAEVEDNTMVTFRNOASRPYFSYSLISYEEDORQGA 1800
 QY 1820 EPRKNFVKPNETKIYFKVQVQHMAPTDEFPCKAWAYFSDVDLEKDHSGLIGPLLVCHT 1879
 DB 1801 EPRKNFVKPNETKIYFKVQVQHMAPTDEFPCKAWAYFSDVDLEKDHSGLIGPLLVCHT 1860
 QY 1880 NTLPAHAGROVTVCEPFFIPIETKSWYTEMNERNCRAPCNQOMEDPTFKENYRPHA 1939
 DB 1861 NTLPAHAGROVTVCEPFFIPIETKSWYTEMNERNCRAPCNQOMEDPTFKENYRPHA 1920
 QY 1940 INGYIMDTPLGLVNAQDQIRIYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1999
 DB 1921 INGYIMDTPLGLVNAQDQIRIYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1980
 QY 2000 GVFEIVEMLPKAGIWRVECLIGELHAGMSTLFLVSNKQOTPLGMASGHIRPOITAS 2059
 DB 1981 GVFEIVEMLPKAGIWRVECLIGELHAGMSTLFLVSNKQOTPLGMASGHIRPOITAS 2040
 QY 2060 GOYQOWAPLARLEYSGSINAWSTKEPSPWIKVDLLAPMI IHGKTQCAROKESLSYISQ 2119
 DB 2041 GOYQOWAPLARLEYSGSINAWSTKEPSPWIKVDLLAPMI IHGKTQCAROKESLSYISQ 2100
 QY 2120 FIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSSGIKHINFPPIIARYIRLHPHTYSIRS 2179
 DB 2101 FIIMYSLDGKKWQYRGNSTGTLMVFFGNVDSSGIKHINFPPIIARYIRLHPHTYSIRS 2160
 QY 2180 TLRVEMLCDDLNSCMLPLGHESKAI SDAQITASSYFTNMFTWSPSKARLHQRSNAWR 2239
 DB 2161 TLRVEMLCDDLNSCMLPLGHESKAI SDAQITASSYFTNMFTWSPSKARLHQRSNAWR 2220
 QY 2240 PQVANNPEKWLQVDQKTMKVTCVTQTQVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
 DB 2221 PQVANNPEKWLQVDQKTMKVTCVTQTQVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
 QY 2300 XVFOGQDSFTFVNVNSLDPPLLTRYLRIHPQSWHYQIALRMEVLGSCBAQDLY 2351

Db 2261 KVFQGNDSPTVNVNSLDPPLLTRYLRHPQSWVHQIALRNEVLGCEAODLY 2332

RESULT 10

US-08-670-707A-2
 ; Sequence 2, Application US/08670707A
 ; Patent No. 5859204
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S.
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 ; STREET: 5370 Manhattan Circle Suite 201
 ; CITY: Boulder
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/670,707A
 ; FILING DATE: 26-JUN-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/US94/13200
 ; FILING DATE: 15-NOV-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/212,133
 ; FILING DATE: 11-MAR-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/864,004
 ; FILING DATE: 07-APR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Greenlee, Lorraine L.
 ; REGISTRATION NUMBER: 27,894
 ; REFERENCE/DOCKET NUMBER: 75-95F
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 303/499-8080
 ; TELEFAX: 303/499-8089
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2332 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: N-terminal
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; TISSUE TYPE: Liver
 ; US-08-670-707A-2

Query Match 99.0%; Score 12293; DB 2; Length 2332;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYILGAVELSDWYMQSDLGSLPVDARPPVPKSPFPNTSVVYKTLFVFTDHLFN 79
 Db 1 ATRRYILGAVELSDWYMQSDLGSLPVDARPPVPKSPFPNTSVVYKTLFVFTVLEFN 60
 QY 80 IAKRPWMLGGLTQAEYVDVTWITLKNWASHPVSLHAGVSYWKASGAEYDDQTSQ 139
 Db 61 IAKRPWMLGGLTQAEYVDVTWITLKNWASHPVSLHAGVSYWKASGAEYDDQTSQ 120
 QY 140 REKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 199

Db 121 REKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180
 QY 200 EGS�AKEKTQTLLKCTLLFAVFEDEGKSWHSEFKNSLMODRDAASAPAWKMTVNGYVNR 259
 Db 181 EGS�AKEKTQTLLKCTLLFAVFEDEGKSWHSEFKNSLMODRDAASAPAWKMTVNGYVNR 240
 QY 260 SLPLGIGCHRKSVYWHVIGWGTTPVHSIFLECHTFLVNRHQASLEISPIFLTAQTLL 319
 Db 241 SLPLGIGCHRKSVYWHVIGWGTTPVHSIFLECHTFLVNRHQASLEISPIFLTAQTLL 300
 QY 320 MDLQFLLSCHISSHQDQWEAYKVDSCEPPEQLRMKNNEAEYDDDLTDSMDVVRP 379
 Db 301 MDLQFLLSCHISSHQDQWEAYKVDSCEPPEQLRMKNNEAEYDDDLTDSMDVVRP 360
 QY 380 DDNSPSFIQIRSVAKKPKTWVHYIAABEEDWDYAPLAPDDRSYKSOYLNNQPORG 439
 Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAABEEDWDYAPLAPDDRSYKSOYLNNQPORG 420
 QY 440 RYKVKVRFMAYTDETFKTEAICHESGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGI 499
 Db 421 RYKVKVRFMAYTDETFKTEAICHESGILGPLLYGEVGTLLIIFKNQASRPYNIYPHGI 480
 QY 500 TDVRPLYSRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGPTKSDPRCLTRYSSVWME 559
 Db 481 TDVRPLYSRRLPKGVKHLKDFILPGEIPKYKWTVTVEDGPTKSDPRCLTRYSSVWME 540
 QY 560 RDLASGLIGPLLI CYKESVDORGNQIMSKRVILFSVFDENRSWYLTENIQFJLNPAG 619
 Db 541 RDLASGLIGPLLI CYKESVDORGNQIMSKRVILFSVFDENRSWYLTENIQFJLNPAG 600
 QY 620 VQLEDPEFOASINMHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSVPSGYTFKH 679
 Db 601 VQLEDPEFOASINMHSINGYVFDLSQLSVCLHEVAYWYILSIGAOTDPLSVPSGYTFKH 660
 QY 680 KVVYEDTLPLPESGETVMSMENPGLWILGCHNSDFRNKGMTALLKVSSCDKNTGDYVE 739
 Db 661 KVVYEDTLPLPESGETVMSMENPGLWILGCHNSDFRNKGMTALLKVSSCDKNTGDYVE 720
 QY 740 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQFNATTIPENDIEKTDPMFAHRTMP 799
 Db 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQFNATTIPENDIEKTDPMFAHRTMP 780
 QY 800 KIQNVSSSDLLMLRQSPPHGLSLSDLOEAKYETPSDDPPSCAIDSNNSLSEMTFRPQ 859
 Db 781 KIQNVSSSDLLMLRQSPPHGLSLSDLOEAKYETPSDDPPSCAIDSNNSLSEMTFRPQ 840
 QY 860 LHSFGDMVFTPSGLQLRNEKLGTTAATLKLDEKSVSTNNLITSTPSDNLAAGTGN 919
 Db 841 LHSFGDMVFTPSGLQLRNEKLGTTAATLKLDEKSVSTNNLITSTPSDNLAAGTGN 900
 QY 920 TSSLGPPSPMEVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENNDKLLBSGLMNSQESSW 979
 Db 901 TSSLGPPSPMEVHYDSOLDTTLFGKXSSPLTESGGPLSLSEENNDKLLBSGLMNSQESSW 960
 QY 980 GKNVSTESGRILFKGKRAHGPALLTKDNALFKVSISSLTKNTKTSNNSATNRKTHIDGPSL 1039
 Db 961 GKNVSTESGRILFKGKRAHGPALLTKDNALFKVSISSLTKNTKTSNNSATNRKTHIDGPSL 1020
 QY 1040 LIENSEPSWQNILESDETEPKKVTPLIHDRMLDKNATLRLNHNMSKNTSSKNMEVQOK 1099
 Db 1021 LIENSEPSWQNILESDETEPKKVTPLIHDRMLDKNATLRLNHNMSKNTSSKNMEVQOK 1080
 QY 1100 KSGPPIPPDAQNDMDGFFKMLFLPESARWTQRTGKNLSNGSGGSPKQLVSLGPKSVEG 1159
 Db 1081 KSGPPIPPDAQNDMDGFFKMLFLPESARWTQRTGKNLSNGSGGSPKQLVSLGPKSVEG 1140
 QY 1160 QNFLSEKXKVVYKGBFTKDVGLKEMWFPSSNLFTNLNHLHNNTNQEKKIQEEIEK 1219
 Db 1141 QNFLSEKXKVVYKGBFTKDVGLKEMWFPSSNLFTNLNHLHNNTNQEKKIQEEIEK 1200
 QY 1220 KETLQENNVVLQIHTVTGCTKMFKNLFLLSRQNVGSGYDGAAYAPVLDQFRSLNDSTNR 1279
 Db 1201 KETLQENNVVLQIHTVTGCTKMFKNLFLLSRQNVGSGYDGAAYAPVLDQFRSLNDSTNR 1260

1280 TKKTHAHSKKEBENLEGLNCTQKQVEKYACTTRISNTSQONFVTPORSKALQKREL 1339
Db 1261 TKKTHAHSKKEBENLEGLNCTQKQVEKYACTTRISNTSQONFVTPORSKALQKREL 1320
Qy 1340 PLEETEELKRIIIVDTSTOWSKMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
Db 1321 PLEETEELKRIIIVDTSTOWSKMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
Qy 1400 POANSSPIAKVSPSPSR2YLRVLPQDNSSHLPAASYKEDSGVQESSHFLQGAKK 1459
Db 1381 POANSSPIAKVSPSPSR2YLRVLPQDNSSHLPAASYKEDSGVQESSHFLQGAKK 1440
Qy 1460 NNLSLAILTLEMTGQREVGSLGTSAATNSVTYKVENVTLPKPDLPKSGKVELLPKVHI 1519
Db 1441 NNLSLAILTLEMTGQREVGSLGTSAATNSVTYKVENVTLPKPDLPKSGKVELLPKVHI 1500
Qy 1520 YOKDLPPTETNSGSHLDLVEGSLQGTGEGAIKWNEANRPKVPFLRVATESSAKTFSK 1579
Db 1501 YOKDLPPTETNSGSHLDLVEGSLQGTGEGAIKWNEANRPKVPFLRVATESSAKTFSK 1560
Qy 1580 LLDPLAWDNHYGTQIPKEWKSQESPKTAPKXDTILSNACSNHAIAINEGQKNP 1639
Db 1561 LLDPLAWDNHYGTQIPKEWKSQESPKTAPKXDTILSNACSNHAIAINEGQKNP 1620
Qy 1640 EIEVTWAKQGRTERLCSONPPVLRKHQREIBITRTTLQSQDEEIDYDDTISVENMKKEDFDIY 1699
Db 1621 EIEVTWAKQGRTERLCSONPPVLRKHQREIBITRTTLQSQDEEIDYDDTISVENMKKEDFDIY 1680
Qy 1700 DEDENQSRSPQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVPOFKVPOEFTD 1759
Db 1681 DEDENQSRSPQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVPOFKVPOEFTD 1740
Qy 1780 CSFTOPLYRGELNEHLGLGPYIRAEVDENIMVTFRNOASRPSYSLSIYVEEDOROGA 1819
Db 1761 CSFTOPLYRGELNEHLGLGPYIRAEVDENIMVTFRNOASRPSYSLSIYVEEDOROGA 1800
Qy 1860 EPRKNFVKNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1879
Db 1801 EPRKNFVKNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVLEKDVHSGLIGPLLVCHT 1860
Qy 1880 NTLNPAHGRQVTQVFALFFITFDTKWYFTENMERNCRAPCNQMEDPTFKENYRHA 1939
Db 1861 NTLNPAHGRQVTQVFALFFITFDTKWYFTENMERNCRAPCNQMEDPTFKENYRHA 1920
Qy 1940 INGYIMDTLPLGLVMAQDQRIKRYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1999
Db 1921 INGYIMDTLPLGLVMAQDQRIKRYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1980
Qy 2000 GVFTVEMLPSKAGTWRVECLIGEHLHAGMSTFLVYGNKQCTPLGMASGHIRDFOITAS 2059
Db 1981 GVFTVEMLPSKAGTWRVECLIGEHLHAGMSTFLVYGNKQCTPLGMASGHIRDFOITAS 2040
Qy 2060 COYGOWAPKLARLHYSGSNASTKEPSWIKVDLLAPMIHGIKTQGAROKFSSLYISQ 2119
Db 2041 COYGOWAPKLARLHYSGSNASTKEPSWIKVDLLAPMIHGIKTQGAROKFSSLYISQ 2100
Qy 2120 FIIMYSLDGKKWQYRGNSTGLTVFFGNVDSSGKIKNIENPPIIARVIRLHPHYSIRS 2179
Db 2101 FIIMYSLDGKKWQYRGNSTGLTVFFGNVDSSGKIKNIENPPIIARVIRLHPHYSIRS 2160
Qy 2180 TIRMLMGCDLNSCMPGLMESKA-SDAQITASSYFTNNFATWSPSKARLHQGRSNAWR 2239
Db 2161 TIRMLMGCDLNSCMPGLMESKA-SDAQITASSYFTNNFATWSPSKARLHQGRSNAWR 2220
Qy 2240 PQVNNPKEWLQVDFQTKMKVGTVTTCQYKSLTSMYVKEFLISSQDGHQWTLFQNGKV 2299
Db 2221 PQVNNPKEWLQVDFQTKMKVGTVTTCQYKSLTSMYVKEFLISSQDGHQWTLFQNGKV 2280
Qy 2300 KVFQGNQDSFTPVVNSLDPPLLTIRLIRHPQSWHQAIALRVEVLGCEAQDLY 2351
Db 2281 KVFQGNQDSFTPVVNSLDPPLLTIRLIRHPQSWHQAIALRVEVLGCEAQDLY 2332

RESULT 11
JS-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6150371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; JS-09-037-601-2

Query Match 99.0%; Score 12293; DB 3; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 20 ATRRYLGAVELSDYMQSDLGELPVDARPPRPVKSPFNNTSVVYKKTLEFVEFTDHLFN 79
Db 1 ATRRYLGAVELSDYMQSDLGELPVDARPPRPVKSPFNNTSVVYKKTLEFVEFTDHLFN 60
Qy 80 IAKRPPFMGLGPTIQAEVYDVTWITLKNMASHFVSLHVGVSYWKASGAEYDDQTSQ 139
Db 61 IAKRPPFMGLGPTIQAEVYDVTWITLKNMASHFVSLHVGVSYWKASGAEYDDQTSQ 120
Qy 140 REKEDDKVFPGSGSHYVQVVKENGPMASDPLCITYSLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFPGSGSHYVQVVKENGPMASDPLCITYSLSHVDLVKDLNSGLIGALLVCR 180
Qy 200 EGSIAKEXTQTLHKFILLFAVFDGSKSWHSETKNSLMQDRDASARAWPKMHTVNGYVNR 259

181 EGS LAKEKTQTLHKFILLFAVDEGKSWSETKNSLMQORDAASAPAWKHWITUNGYNR 240
260 SLPGILGCHKRSVYWHVIMGWTTPFVHSIFLBCHTFLVNRHQASLEISPIFLTAQTL 319
241 SLPGILGCHKRSVYWHVIMGWTTPFVHSIFLBCHTFLVNRHQASLEISPIFLTAQTL 300
320 MDLQGLLCHTSSHOHDMGEMAVKVDSCPEEPQLRMKNNEBAEDYDDDLTDSMDVVR 379
301 MDLQGLLCHTSSHOHDMGEMAVKVDSCPEEPQLRMKNNEBAEDYDDDLTDSMDVVR 360
380 DDNSPSFTQIRSVAKKHPTWVHYIAAEEBWDVAPLVADDDRSYKSOYLNNQGRIG 439
361 DDNSPSFTQIRSVAKKHPTWVHYIAAEEBWDVAPLVADDDRSYKSOYLNNQGRIG 420
440 RYKVKVRFMAYTDEFTKREAOHESGILGPLYGEVGTLLIIFKNQASRPVNIYPHGI 499
421 RYKVKVRFMAYTDEFTKREAOHESGILGPLYGEVGTLLIIFKNQASRPVNIYPHGI 480
500 TDVRPLYSRRLPKGVKHLXDFPILPGEI PKYKWTVTVEGPTKSDPRCLTRYYSFVWNE 559
481 TDVRPLYSRRLPKGVKHLXDFPILPGEI PKYKWTVTVEGPTKSDPRCLTRYYSFVWNE 540
560 RLASGLIPLIICYKESVDQRNQIMSDKRNWILFSVFDENRSWYLTENIQRFUPNPAG 619
541 RLASGLIPLIICYKESVDQRNQIMSDKRNWILFSVFDENRSWYLTENIQRFUPNPAG 600
620 VOLEDEFOASNIHMSINGVFDLSQLSVCLHEVAYWYILSICAOTDFLSVFPSPGVTXKH 679
601 VOLEDEFOASNIHMSINGVFDLSQLSVCLHEVAYWYILSICAOTDFLSVFPSPGVTXKH 660
680 XWYEDTLTLPFSGETVPMSENPGWLILGCHNSDFRNKGMTALLKVSXCDKNTGDYVE 739
661 XWYEDTLTLPFSGETVPMSENPGWLILGCHNSDFRNKGMTALLKVSXCDKNTGDYVE 720
740 DSYEDISAYLLSKNNAIBRSFSQNSRHSTQKQFNATTIPENDIEKTDWFAHRTMP 799
721 DSYEDISAYLLSKNNAIBRSFSQNSRHSTQKQFNATTIPENDIEKTDWFAHRTMP 780
800 KTVNSSSDLLMLLROSPTPHGLSLSDLOBAKYETFSDDPSGAISSNNSLSEMTXFRPQ 859
781 KTVNSSSDLLMLLROSPTPHGLSLSDLOBAKYETFSDDPSGAISSNNSLSEMTXFRPQ 840
860 LHSXGDMVFTPSGILRLNEKLTGTAATLKLDPKVSSTNNLIISTIPSDNLAAGTGN 919
841 LHSXGDMVFTPSGILRLNEKLTGTAATLKLDPKVSSTNNLIISTIPSDNLAAGTGN 900
920 TSSLGPPSPMPVHYDSOLDTTLFGKXSSPLTESGGPLSLSENNDSKLLSGLMNOQESSW 979
901 TSSLGPPSPMPVHYDSOLDTTLFGKXSSPLTESGGPLSLSENNDSKLLSGLMNOQESSW 960
980 GKNVSTESGRLLFKGKRAHGFALLTKDNALFKVISLSLLKTKNTKNSNATNRKTHIDGSL 1039
961 GKNVSTESGRLLFKGKRAHGFALLTKDNALFKVISLSLLKTKNTKNSNATNRKTHIDGSL 1020
1040 LIENSFSVWQNLIESDTEPKVTPLIHDRMLMDKNATLRLNHNKNTTSSKNMVMQOK 1099
1021 LIENSFSVWQNLIESDTEPKVTPLIHDRMLMDKNATLRLNHNKNTTSSKNMVMQOK 1080
1100 KEGPIPPDAQNPDMPFKMLFLPESARWIQRTGKNSLNSGQSPKQIVSIPGPKSVEG 1159
1081 KEGPIPPDAQNPDMPFKMLFLPESARWIQRTGKNSLNSGQSPKQIVSIPGPKSVEG 1140
1160 QNFLEKKNVYVKGGEFTKDVGLKEMVFPSSRNLFNLNLFENNTHNQEKKIQBEIEK 1219
1141 QNFLEKKNVYVKGGEFTKDVGLKEMVFPSSRNLFNLNLFENNTHNQEKKIQBEIEK 1200
1220 KSTLQENWVLPQIHVTGTQKPMKNLPLLSTRQNVVEGSDGAYAPVLQDFSLNDSTNR 1279
1201 KSTLQENWVLPQIHVTGTQKPMKNLPLLSTRQNVVEGSDGAYAPVLQDFSLNDSTNR 1260
1280 TKKHTAHFSKKGEEENLEGLNQTQKQIIVEKYACTRISNTSOQNFVTOXSRLAKQFRL 1339

1261 TKKHTAHFSKKGEEENLEGLNQTQKQIIVEKYACTRISNTSOQNFVTOXSRLAKQFRL 1320
1340 P-EBTELKRIIIVDDTSTQMSXMKHLT2STILTQDYNEKEKGAITOSPLSCLTRSHSI 1399
1321 P-EBTELKRIIIVDDTSTQMSXMKHLT2STILTQDYNEKEKGAITOSPLSCLTRSHSI 1380
1400 POANRSPPIAKVSSFPPIRPIYLRVLFDQNSSHLPAA5YRKDQSGVQESSHFLQGAKK 1459
1381 POANRSPPIAKVSSFPPIRPIYLRVLFDQNSSHLPAA5YRKDQSGVQESSHFLQGAKK 1440
1460 NNL5LAITLLEMTGQREVGLSGTSA5YVYKXKVENTVLPKDLPKTSKGVKELLIPKVI 1519
1441 NNL5LAITLLEMTGQREVGLSGTSA5YVYKXKVENTVLPKDLPKTSKGVKELLIPKVI 1500
1520 YOKDIPPTETNGSPGHLDLVEGSLLOGTGEGA1KWEANRPGKVPFLRVATESSAKTFSK 1579
1501 YOKDIPPTETNGSPGHLDLVEGSLLOGTGEGA1KWEANRPGKVPFLRVATESSAKTFSK 1560
1580 LLDPLAWNHNHYGTQIPKEEWSQESPKTAPKXKDTILSLNACSNHAI1A1NAINEGONKP 1639
1561 LLDPLAWNHNHYGTQIPKEEWSQESPKTAPKXKDTILSLNACSNHAI1A1NAINEGONKP 1620
1640 EIEVTAQKQRTPELCSQNPVPLKHOREITRTTLQSQOEEIDYDDTISVEMKEDFDIY 1699
1621 EIEVTAQKQRTPELCSQNPVPLKHOREITRTTLQSQOEEIDYDDTISVEMKEDFDIY 1680
1700 DEDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVPQKVVQFERTD 1759
1681 DEDENOSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVPQKVVQFERTD 1740
1760 GSFTQPLYRGELNEHLGLLGPYIRAEVEDCNMTWTFRNOASRPYSYSSLSIYEEQROGA 1819
1741 GSFTQPLYRGELNEHLGLLGPYIRAEVEDCNMTWTFRNOASRPYSYSSLSIYEEQROGA 1800
1820 EPRKXVFNPKETKTYFWKVOHMAPTKDBFPCKAWAYFSDVDLEKDVHSGILGPLLVCHT 1879
1801 EPRKXVFNPKETKTYFWKVOHMAPTKDBFPCKAWAYFSDVDLEKDVHSGILGPLLVCHT 1860
1880 NTLNPAHGRQVTVQSFALPFTTIFDFTKSWYFTENNERNCRAPCNIQMEDPFTKENVRPHA 1939
1861 NTLNPAHGRQVTVQSFALPFTTIFDFTKSWYFTENNERNCRAPCNIQMEDPFTKENVRPHA 1920
1940 INGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVYRKEEYKVALYNLYP 1999
1921 INGYIMDTLPLGVMAQDQIRWYLLSMGNSNENIHSIHFSGHVFTVYRKEEYKVALYNLYP 1980
2000 GVFETVEMLPKAGIWRVBCUIGELHAGMSTLFLVYGNKQOTPLCMASGHIRDQITAS 2059
1981 GVFETVEMLPKAGIWRVBCUIGELHAGMSTLFLVYGNKQOTPLCMASGHIRDQITAS 2040
2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIJHGKTCQARQKFSLSYTSQ 2119
2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIJHGKTCQARQKFSLSYTSQ 2100
2120 FIIMYSLDCKKQWTVRGNSTGTLVFFGNVDSSG1KHNI FNPPIIARIYLRHPTHYSIRS 2179
2101 FIIMYSLDCKKQWTVRGNSTGTLVFFGNVDSSG1KHNI FNPPIIARIYLRHPTHYSIRS 2160
2180 TLRMELMGCDLNSCMLGMSKALSDAQITASSYFTNMFPATWSKARLHLCQSNAWR 2239
2161 TLRMELMGCDLNSCMLGMSKALSDAQITASSYFTNMFPATWSKARLHLCQSNAWR 2220
2240 PQVNPKEWLQVDFOKTKMKTGVTTCQGVKSLLTSMYKFEPLISSQDGHOWTLFPQNGKV 2299
2221 PQVNPKEWLQVDFOKTKMKTGVTTCQGVKSLLTSMYKFEPLISSQDGHOWTLFPQNGKV 2280
2300 KVFQGNQDSFTPVVNSLDPPLLTLYRLRHPOSVWHQIALRVEVLGCEAQDLY 2351
2281 KVFQGNQDSFTPVVNSLDPPLLTLYRLRHPOSVWHQIALRVEVLGCEAQDLY 2332


```

; Sequence 2, Application US/09315179
; Patent No. 6376463
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VII:
; FILE REFERENCE: 75-95H
; CURRENT APPLICATION NUMBER: US/09/315,179
; EARLIER FILING DATE: 1999-05-20
; EARLIER APPLICATION NUMBER: U.S. 09/037,601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: U.S. 08/670,707
; EARLIER FILING DATE: 1996-06-26
; EARLIER APPLICATION NUMBER: PCT/US97/11155
; EARLIER FILING DATE: 1997-06-26
; EARLIER APPLICATION NUMBER: PCT/US94/13200
; EARLIER FILING DATE: 1994-11-15
; EARLIER APPLICATION NUMBER: U.S. 08/212,133
; EARLIER FILING DATE: 1994-03-11
; EARLIER APPLICATION NUMBER: U.S. 07/864,004
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TIPS: PRT
; ORGANISM: Homo sapiens
; US-09-315-179-2

Query Match          99.0%; Score 12293; DB 4; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYYLGAVELSDWYMQSDLGELPVDARPPRPVPSPPFNVSVVYKTLFVEFTDHLFN 79
DB 1 ATRRYYLGAVELSDWYMQSDLGELPVDARPPRPVPSPPFNVSVVYKTLFVEFTVHLEN 60
QY 80 IAKPRPPWMLLGPITQAEVYDVTWITLKNMASHPVSLHAGVSVKASGAEYDDQTSQ 139
DB 61 IAKPRPPWMLLGPITQAEVYDVTWITLKNMASHPVSLHAGVSVKASGAEYDDQTSQ 120
QY 140 REKEDKVPFGGSHYVVOVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDKVPFGGSHYVVOVLKENGPMASDPLCLTYSLYSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTQTLKFTLLPAVDEGKSHSETKNSLMODRDAASARAWPKVHTVANGYNR 259
DB 181 EGS LAKEKTQTLKFTLLPAVDEGKSHSETKNSLMODRDAASARAWPKVHTVANGYNR 240
QY 260 SLPGILGCHRKSVYVHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 319
DB 241 SLPGILGCHRKSVYVHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300
QY 320 MDLGQFLFCHLSSHQHDMGEAYVVDSCPEPQLRMKMKNBEAEDYDDDLTDEMDVVR 379
DB 301 MDLGQFLFCHLSSHQHDMGEAYVVDSCPEPQLRMKMKNBEAEDYDDDLTDEMDVVR 360
QY 380 DDNSPSTQIRSVAKHPKTHVHYIAAEEEDWDYAPLVAPDDRSYKSYQLNNGPQRIG 439
DB 361 DDNSPSTQIRSVAKHPKTHVHYIAAEEEDWDYAPLVAPDDRSYKSYQLNNGPQRIG 420
QY 440 RYKAVRMAVYDTEFKTREAOHESGILGPLYGEVGDTHLIIIPKQASRPYNTYPHGI 499
DB 421 RYKAVRMAVYDTEFKTREAOHESGILGPLYGEVGDTHLIIIPKQASRPYNTYPHGI 480
QY 500 TDVRLYRRRLPKGVKHLKDPILPGEIFKYKWTVTVDGPTKSDPRCLTRYSSFVWME 559
DB 481 TDVRLYRRRLPKGVKHLKDPILPGEIFKYKWTVTVDGPTKSDPRCLTRYSSFVWME 540
QY 560 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPAG 619
DB 541 RDLASGLIGPLLCYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQRFPLNPAG 600
QY 620 VQLEDPPEFQASNMHMSINGYVFDLSQLSVCLHEVAYWILSIGAQTDFLSVFFSGYTFKH 679

```

Db 1681 DEDNQSPRSFQKTRHYFIAAVERLWDYGMSSSPHLNRNQAQSGSVQPKKVVQFEFTD 1740
 Qy 1760 GSFTOPLVGEINELHGLLGPVIRAEVDNIMVTNRNQAASRYSYSSLSIYEEEDORQA 1819
 Db 1741 GSFTOPLVGEINELHGLLGPVIRAEVDNIMVTNRNQAASRYSYSSLSIYEEEDORQA 1800
 Qy 1820 EPRKNFVAPNETKTYFWKVQHMAPTKDFCDKAWAYFSDVDLEKDVHSGGLICPLLVCHT 1879
 Db 1801 EPRKNFVAPNETKTYFWKVQHMAPTKDFCDKAWAYFSDVDLEKDVHSGGLICPLLVCHT 1860
 Qy 1880 NTINPAHGRQVTVQSPALPFTTIDETKSWYFTENMERNCRAPCNTQMEDPTPKENYRPHA 1939
 Db 1861 NTINPAHGRQVTVQSPALPFTTIDETKSWYFTENMERNCRAPCNTQMEDPTPKENYRPHA 1920
 Qy 1940 INGYINDTLPGLVMAQDQIRWVLLSMGSENIHSHRSGHVFTVRKKEEYKMAVLYNLYP 1999
 Db 1921 INGYINDTLPGLVMAQDQIRWVLLSMGSENIHSHRSGHVFTVRKKEEYKMAVLYNLYP 1980
 Qy 2000 GVFETVEXPLPSKAGIWRVCECLIGELHAGMSTLFLVYSNKCOTPLGMASGHIRDQITAS 2059
 Db 1981 GVFETVEXPLPSKAGIWRVCECLIGELHAGMSTLFLVYSNKCOTPLGMASGHIRDQITAS 2040
 Qy 2060 GYGQWAPKLAHLHSGSINASTKEPFSWIKVDLLAPMIHGIKTQGARQKPFSSLYISQ 2119
 Db 2041 GYGQWAPKLAHLHSGSINASTKEPFSWIKVDLLAPMIHGIKTQGARQKPFSSLYISQ 2100
 Qy 2120 FIIMYSLDGKQWQYRGNGSTGLMVPFGNVDSGGIKHNFNPPIIARYIRLHPHTYSIRS 2179
 Db 2101 FIIMYSLDGKQWQYRGNGSTGLMVPFGNVDSGGIKHNFNPPIIARYIRLHPHTYSIRS 2160
 Qy 2180 TLRMELMGCDLNSCSMPLQWESKAISDAQITASVYFTNMFATWSPSKARLHLQGRSNWR 2239
 Db 2161 TLRMELMGCDLNSCSMPLQWESKAISDAQITASVYFTNMFATWSPSKARLHLQGRSNWR 2220
 Qy 2240 PQVNNFKEWLQVDQKTXKVTGTTQGVKSLTSTYVVKSLTSSODGQWTLFPQNGKY 2299
 Db 2221 PQVNNFKEWLQVDQKTXKVTGTTQGVKSLTSTYVVKSLTSSODGQWTLFPQNGKY 2280
 Qy 2300 KYFQGNQDSFTPVNSLDPLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 2351
 Db 2281 KYFQGNQDSFTPVNSLDPLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 2332

RESULT 13

US-09-523-656-2
 ; Sequence 2, Application US/09523656
 ; Patent No. 6458563

GENERAL INFORMATION:

; APPLICANT: Lollar S., John
 ; TITLE OF INVENTION: MODIFIED FACTOR VII
 ; FILE REFERENCE: 75-951
 ; CURRENT APPLICATION NUMBER: US/09/523,656
 ; CURRENT FILING DATE: 2000-03-10
 ; EARLIER APPLICATION NUMBER: 09/037,601
 ; EARLIER FILING DATE: 1998-03-10
 ; EARLIER APPLICATION NUMBER: 08/670,707
 ; EARLIER FILING DATE: 1996-06-26
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2

; LENGTH: 2332

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-523-656-2

Query Match

Best Local Similarity 99.9%; Score 12293; DB 4; Length 2332;

Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 20

ATERYVLGAVELSDYMQSDLGELPVDARPPRPVPSFPNTSWYKTLFVEFTDHLFN 79

Db 1

ATERYVLGAVELSDYMQSDLGELPVDARPPRPVPSFPNTSWYKTLFVEFTDHLFN 60

Qy 80 IAKPRPPMGLGTTIOAEVYDTVTWVILKNWASHPVSLHAVGVSYWKAASGAEYDDQTSQ 139
 Db 61 IAKPRPPMGLGTTIOAEVYDTVTWVILKNWASHPVSLHAVGVSYWKAASGAEYDDQTSQ 120
 Qy 140 REKEDDKVPFGSSTYVQVILKENGPMASDPLCTYSYLSHVLDLWKLNSGLIGALLVCR 199
 Db 121 REKEDDKVPFGSSTYVQVILKENGPMASDPLCTYSYLSHVLDLWKLNSGLIGALLVCR 180
 Qy 200 EGSILAKETQTLHFIILLFAVFDGKSWHSETKSLMODRDAASARAWPKMHTVNGYNR 259
 Db 181 EGSILAKETQTLHFIILLFAVFDGKSWHSETKSLMODRDAASARAWPKMHTVNGYNR 240
 Qy 260 SLPLGLIGCHRXSVMHVIGMGTTPVEVHSIFLEGHTFLVNRHQASLEISPIITFATAQTL 319
 Db 241 SLPLGLIGCHRXSVMHVIGMGTTPVEVHSIFLEGHTFLVNRHQASLEISPIITFATAQTL 300
 Qy 320 MDLGOPILSCHISSHOHDGMEAYVKVDCPEBPQLRMKNNEEADYDDDLTDSMDVVRP 379
 Db 301 MDLGOPILSCHISSHOHDGMEAYVKVDCPEBPQLRMKNNEEADYDDDLTDSMDVVRP 360
 Qy 380 DDNSPSPSIQIRSVAKKHPKTWVHYIAAEBEDWDVAPLVAPDDRSYKSOYLNNPQIRG 439
 Db 361 DDNSPSPSIQIRSVAKKHPKTWVHYIAAEBEDWDVAPLVAPDDRSYKSOYLNNPQIRG 420
 Qy 440 RYKXKVRFMAYTDETFKTREAIQHESGILGPELLYGEVGDTLIIIPNQASRPYNIYPHGI 499
 Db 421 RYKXKVRFMAYTDETFKTREAIQHESGILGPELLYGEVGDTLIIIPNQASRPYNIYPHGI 480
 Qy 500 TDVRLYSRRILPKGVKHLKDPILPGEIFKYKWTVTVEDGGTKSDPRCLTYYSFVNM 559
 Db 481 TDVRLYSRRILPKGVKHLKDPILPGEIFKYKWTVTVEDGGTKSDPRCLTYYSFVNM 540
 Qy 560 RDLASGLIGPILLICVKEVDQGNQIMSDKNVILFSVFDENRSMYLTENTQRFLPNPAG 619
 Db 541 RDLASGLIGPILLICVKEVDQGNQIMSDKNVILFSVFDENRSMYLTENTQRFLPNPAG 600
 Qy 620 VQLEDPFQAGNIMESINGYVFDLSIQLSVCLHEVAYWVILSIAQTDFLSVFSGYTFKH 679
 Db 601 VQLEDPFQAGNIMESINGYVFDLSIQLSVCLHEVAYWVILSIAQTDFLSVFSGYTFKH 660
 Qy 680 KNAVYEDTLTLFPFSGETVMSNENGLATLGHNSDFENRGMTALLKYSSCDKNIGDYVE 739
 Db 661 KNAVYEDTLTLFPFSGETVMSNENGLATLGHNSDFENRGMTALLKYSSCDKNIGDYVE 720
 Qy 740 DSYEDISAYLLSKNAIEPRSPSONSRHPSSTROKQFNATTIPENDIEKTDPEFAHRTMP 799
 Db 721 DSYEDISAYLLSKNAIEPRSPSONSRHPSSTROKQFNATTIPENDIEKTDPEFAHRTMP 780
 Qy 800 KIONVSSDILLMLRQSPTPHGLS-SDIQAKEYTFSDDPSPGATDSNNLSLSEMTHFAPQ 859
 Db 781 KIONVSSDILLMLRQSPTPHGLS-SDIQAKEYTFSDDPSPGATDSNNLSLSEMTHFAPQ 840
 Qy 860 LHSXGDMVFTPESGLOLRNEKLGTTAATLKLDPKVSSTNNLIISTIPSDNLAAGTDN 919
 Db 841 LHSXGDMVFTPESGLOLRNEKLGTTAATLKLDPKVSSTNNLIISTIPSDNLAAGTDN 900
 Qy 920 TSSLGPPMPVHYDSQDLDTTLFGKXSSPLTESGGPLSLSENNDSKLLSGLMNSQSSSW 979
 Db 901 TSSLGPPMPVHYDSQDLDTTLFGKXSSPLTESGGPLSLSENNDSKLLSGLMNSQSSSW 960
 Qy 980 GKNVSTSGRLFKGKRAHGPAALLTKDNALFKVSIKLTNKTNSNATNRKTHTDGSL 1039
 Db 961 GKNVSTSGRLFKGKRAHGPAALLTKDNALFKVSIKLTNKTNSNATNRKTHTDGSL 1020
 Qy 1040 LSENSPSWQNLIESDTFCKVTPLLIHDRMLMDKNATLRNLNHNKNTSSKNMVMQOK 1099
 Db 1021 LSENSPSWQNLIESDTFCKVTPLLIHDRMLMDKNATLRNLNHNKNTSSKNMVMQOK 1080
 Qy 1100 KEGPIPPDAQNPMDFEFKMLFLPESARWIORTHGKNSLNSGQSPKQVLSLGPESKVEG 1159
 Db 1081 KEGPIPPDAQNPMDFEFKMLFLPESARWIORTHGKNSLNSGQSPKQVLSLGPESKVEG 1140
 Qy 1160 QNLFSEKKNVVKGEFTKDVGLKEMVFPSSNLFILTNLDNLHNNTNQEKIOSEIEK 1219

Db 1141 QNFLSEKXKVVVGKEFTKDVCLKEMVFPSSRLPLTNLDLHNNTHNQKXIQEETK 1200
Qy 1220 KETLLIQENAVLPQIHVTVGTRKFMKNLPLSTRQNVESYDAGYAPVLQDPRSLNDSNR 1279
Db 1201 KETLLQENAVLPQIHVTVGTRKFMKNLPLSTRQNVESYDAGYAPVLQDPRSLNDSNR 1260
Qy 1280 TKKHTAHSKKEEENLEGLNQTQKQIIEKACTTRISPNSTQONFVTRQSKRALKQFRL 1339
Db 1261 TKKHTAHSKKEEENLEGLNQTQKQIIEKACTTRISPNSTQONFVTRQSKRALKQFRL 1320
Qy 1340 PLEETELESKRIIVDDTSTQWSXNMKHLTSTLTQIDYNEKKGAIQTSPSDCLTRSHSI 1399
Db 1321 PLEETELESKRIIVDDTSTQWSXNMKHLTSTLTQIDYNEKKGAIQTSPSDCLTRSHSI 1380
Qy 1400 POANSPPIAKVSSPSTRPIYLTRVLFODNSSHLPAASYRKXKDSGVQESSHFLOGAKK 1459
Db 1381 POANSPPIAKVSSPSTRPIYLTRVLFODNSSHLPAASYRKXKDSGVQESSHFLOGAKK 1440
Qy 1460 NNLSLAILTLEMTGQREVGSLGTGATKSVTYKKVENTVLPKPDLPKTSKGVELLPKVHI 1519
Db 1441 NNLSLAILTLEMTGQREVGSLGTGATKSVTYKKVENTVLPKPDLPKTSKGVELLPKVHI 1500
Qy 1520 YOKDLPPTSTNSGPHLDLVESLLOQTTEGAIKWNEANRPKVFPFLVATESSAKTPSK 1579
Db 1501 YOKDLPPTSTNSGPHLDLVESLLOQTTEGAIKWNEANRPKVFPFLVATESSAKTPSK 1560
Qy 1580 LLDPLAWNHVGTQIPKEBWSQKSPKTAPEKDDTILSNACSNHAIINAINEGQNK 1639
Db 1561 LLDPLAWNHVGTQIPKEBWSQKSPKTAPEKDDTILSNACSNHAIINAINEGQNK 1620
Qy 1640 EIEVTWAKQGRTERLCSQNPVVKHQREIRTTLLQSQDEEIDYDDTISVEMKEDFDIY 1699
Db 1621 EIEVTWAKQGRTERLCSQNPVVKHQREIRTTLLQSQDEEIDYDDTISVEMKEDFDIY 1680
Qy 1700 DEDENQSPRSQKTRHYFIAAVERLWBYGMSSSHVLNRAQSGSVQPKKVPQEBTD 1759
Db 1681 DEDENQSPRSQKTRHYFIAAVERLWBYGMSSSHVLNRAQSGSVQPKKVPQEBTD 1740
Qy 1760 GSFTQPLRGELNEHLGLGPVIRAEVDNIMVTRNOASRPYSFVSSLSIYEEDQROGA 1819
Db 1741 GSFTQPLRGELNEHLGLGPVIRAEVDNIMVTRNOASRPYSFVSSLSIYEEDQROGA 1800
Qy 1820 BPRKNFVXPNETKTYFWKQVHMAPTKDEFCDAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
Db 1801 BPRKNFVXPNETKTYFWKQVHMAPTKDEFCDAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
Qy 1880 NTLNPAHGRQVTVQEPALPFTIFDETQSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQEPALPFTIFDETQSWYFTENMERNCRAPCNQMEDPTFKENYRPHA 1920
Qy 1940 INGYTMDTLPGLVMAQDQRIRWYLLSMGSNENIHFSGHVFTVRKKEEYKXALYNLYP 1999
Db 1921 INGYTMDTLPGLVMAQDQRIRWYLLSMGSNENIHFSGHVFTVRKKEEYKXALYNLYP 1980
Qy 2000 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQOTPLGMASGHIRDFOITAS 2059
Db 1981 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKQOTPLGMASGHIRDFOITAS 2040
Qy 2060 GOYGQWAPKLARLHYSGSNANSTKEPFWIKVDLLAPMIHGIKTQAGARQKQFSSLYISQ 2119
Db 2041 GOYGQWAPKLARLHYSGSNANSTKEPFWIKVDLLAPMIHGIKTQAGARQKQFSSLYISQ 2100
Qy 2120 FLIYSLDGKQKQYRGNSTGLTVFFVGNVDSG:KHNI FNPPIIARYIRLPHTHYSIRS 2179
Db 2101 FLIYSLDGKQKQYRGNSTGLTVFFVGNVDSG:KHNI FNPPIIARYIRLPHTHYSIRS 2160
Qy 2180 TURBELMGDLNCSNPLGWSKALSDAQITASSYFTNNFATWSPSKARLHLQGRSNARW 2239
Db 2161 TURBELMGDLNCSNPLGWSKALSDAQITASSYFTNNFATWSPSKARLHLQGRSNARW 2220
Qy 2240 POWNNEKEWLQVDFQKTYKVTGTTQGVKSLITSYVVEFLISSQDGHQWTLFFQNGKV 2299

Db 2221 PQVANNPEWLQVDFQKTKMTGVTGTTQGVKSLITSYVVEFLISSQDGHQWTLFFQNGKV 2280
Qy 2300 KVFQGNQDSFTPVVNSDDPPLLTLYLRHPOSVWHQIALRMEVLGCEAODLY 2351
Db 2281 KVFQGNQDSFTPVVNSDDPPLLTLYLRHPOSVWHQIALRMEVLGCEAODLY 2332
RESULT 14
PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03275
; FILING DATE: 19930407
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU 106PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
PCT-US93-03275-4
Query Match 95.0%; Score 12293; DB 5; Length 2332;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 20 ATRYYLGAVELSWDYMOSDGLGELPVDARFPFRPVKSPFPFNSTVYVYKTLFVETDHLFN 79
Db 1 ATRYYLGAVELSWDYMOSDGLGELPVDARFPFRPVKSPFPFNSTVYVYKTLFVETDHLFN 60
Qy 80 IAKRPPPMWGLGPTIQAEVYDTVTITLKNASHPVSLHAGVSVYKASGEAYDDQTSQ 139
Db 61 IAKRPPPMWGLGPTIQAEVYDTVTITLKNASHPVSLHAGVSVYKASGEAYDDQTSQ 120
Qy 140 REKEDDKVFPQGSSTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 199
Db 121 REKEDDKVFPQGSSTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 180
Qy 200 EGSIAKETQTLHKFILLFAVFDKESWHSSTKXSLMODRDAASARAPKMTVNGYNR 259

181 EGSUAKXTQTLKHLFAVDEGSKHSEIKNSLMQDROASARAWPKHVTNGYVR 240
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabst, Patricia L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; PCT-US94-13200-2

Query Match 99.0%; Score 12293; DB 5; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYILGAVELSDWYMQSDLGELPVDARFPVRPKSPFPNTSVVYKTLFVEFTDHLFN 79
DB 1 ATRRYILGAVELSDWYMQSDLGELPVDARFPVRPKSPFPNTSVVYKTLFVEFTVHLFN 60

QY 80 IAKPRPPWMLLGPTIOAEVYTVITLKNMASHPVSLHAVGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPWMLLGPTIOAEVYTVITLKNMASHPVSLHAVGVSYWKASGAEYDDQTSQ 120

QY 140 RKEDEDKVPPGSGSHYVWQVLKENGEMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
DB 121 RKEDEDKVPPGSGSHYVWQVLKENGEMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180

QY 200 EGS LAKEKTQTLHKILLZAVFDGKSWHSETKNSLMQDRDAASAPAWPKMETVNGYVNR 259
DB 181 EGS LAKEKTQTLHKILLZAVFDGKSWHSETKNSLMQDRDAASAPAWPKMETVNGYVNR 240

QY 260 SLPGLIGCRKSVYWHVIGWTTPEVHSIFLEGHTFLVNRHQASLEISPTFLTAQTL 319
DB 241 SLPGLIGCRKSVYWHVIGWTTPEVHSIFLEGHTFLVNRHQASLEISPTFLTAQTL 300

QY 320 MDLGFLLSCHISSHQHDGMEAVVVDSCPEEPQLRMKNNEAEADYDDDLTDSMDVVR 379
DB 301 MDLGFLLSCHISSHQHDGMEAVVVDSCPEEPQLRMKNNEAEADYDDDLTDSMDVVR 360

QY 380 DDNSPSPTQIRSAVKHPKTVWHYIAAEEBWDYAPLVAPDDRYSYKQYLNNGPQIRG 439

```

QY 1520 YKDLPTTTSNGSPGLDINEGSLQGTGEGALKWNEANRPGKVPFLRVATSSAKTPSK 1579
DB 1501 YKDLPTTTSNGSPGLDINEGSLQGTGEGALKWNEANRPGKVPFLRVATSSAKTPSK 1560
QY 1580 LLDPLAWDNHYGTQIPKEWKSQEKSPKTAFFKKOTILSLNACSNHAIAAINEQNKRP 1639
DB 1561 LLDPLAWDNHYGTQIPKEWKSQEKSPKTAFFKKOTILSLNACSNHAIAAINEQNKRP 1620
QY 1640 EIEVTVAKQGRIBERLCSQNPVLKRHOREITRTTLOSDBEIDYDDTISVEMKKEDFDIY 1699
DB 1621 EIEVTVAKQGRIBERLCSQNPVLKRHOREITRTTLOSDBEIDYDDTISVEMKKEDFDIY 1680
QY 1700 DEENOSPSFQKKTTHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVPOFKKVVFQEFDT 1759
DB 1681 DEENOSPSFQKKTTHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVPOFKKVVFQEFDT 1740
QY 1760 GSTQPLYRGELNEHLGLGEPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEEDQORQGA 1819
DB 1741 GSTQPLYRGELNEHLGLGEPYIRAEVEDNIMVTFRNQASRPYSYSSLSIYEEDQORQGA 1800
QY 1820 EPRKNVKNETKTYFWKVQHHMADPKDEDDCKAWAYFSDVDLEKDVHSGSLIGPLIVCHT 1879
DB 1801 EPRKNVKNETKTYFWKVQHHMADPKDEDDCKAWAYFSDVDLEKDVHSGSLIGPLIVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALPFTIIDEKTSWYFTENMERNCRAPCNIQMEDPTFKENYRFHA 1939
DB 1861 NTLNPAHGRQVTVQEFALPFTIIDEKTSWYFTENMERNCRAPCNIQMEDPTFKENYRFHA 1920
QY 1940 INGYIMDTLPLGMAQDORIRWLLSGNSNENIHSIHFSGHVFTVKKBEYKXALYNLYP 1999
DB 1921 INGYIMDTLPLGMAQDORIRWLLSGNSNENIHSIHFSGHVFTVKKBEYKXALYNLYP 1980
QY 2000 GVFEVEMLPKAGIWRVECLIGELHAGNSTLFLVYSNKCOTPLGWSAGHIEDFOITAS 2059
DB 1981 GVFEVEMLPKAGIWRVECLIGELHAGNSTLFLVYSNKCOTPLGWSAGHIEDFOITAS 2040
QY 2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPWIIHGIKTQGARQKFTSSLYISQ 2119
DB 2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPWIIHGIKTQGARQKFTSSLYISQ 2100
QY 2120 FIIMYSLDGKKWOTYRGNSTGTLNVFFGNVDSGGIKHINFPPIIARYIRLHPTHYSIRS 2179
DB 2101 FIIMYSLDGKKWOTYRGNSTGTLNVFFGNVDSGGIKHINFPPIIARYIRLHPTHYSIRS 2160
QY 2180 TLRMLMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMPATWSPSKARLHLQGRSNAWR 2239
DB 2161 TLRMLMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMPATWSPSKARLHLQGRSNAWR 2220
QY 2240 PQVNNPKEMI QVDFQKTMKVTVGTTCQVKSLLATSMYVKEFLISSQDGHQWTLFFQNGKY 2299
DB 2221 PQVNNPKEMI QVDFQKTMKVTVGTTCQVKSLLATSMYVKEFLISSQDGHQWTLFFQNGKY 2280
QY 2300 KVFQGNQDSFTPVVNSLDPPLLRYLRIHPQSWWHQIALRMEVLGCEAODLY 2351
DB 2281 KVFQGNQDSFTPVVNSLDPPLLRYLRIHPQSWWHQIALRMEVLGCEAODLY 2332

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:08:32 ; Search time 61.5 Seconds
(without alignments)
10051.498 Million cell updates/sec

Title: NP000123-328

Perfect score: 12416

Sequence: 1 MQIELSTCFCLLRFCFSFSA.....VVHQIALRMVGLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues

Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US07_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12410	100.0	2351	14	US-10-132-829-4
2	12410	100.0	2351	14	US-10-172-712-27
3	12410	100.0	2351	14	US-10-133-907-4
4	12401	99.9	2351	12	US-10-411-037-30
5	12401	99.9	2351	12	US-10-411-026-30
6	12306	99.1	2332	15	US-10-360-101-229
7	12293	99.0	2332	9	US-09-957-641-2
8	12293	99.0	2332	14	US-10-187-319-2
9	12293	99.0	2332	14	US-10-131-510A-2
10	12293	99.0	2332	15	US-10-445-235-2
11	12293	99.0	2332	16	US-10-239-498A-2
12	8827	71.1	2319	14	US-10-187-319-5
13	8827	71.1	2319	14	US-10-131-510A-6
14	8668	65.0	2133	14	US-10-187-319-37
15	8668	65.0	2133	14	US-10-131-510A-37

16	7404	59.6	1471	12	US-10-681-970-2	Sequence 2, Appli
17	7404	59.6	1471	13	US-10-095-718-2	Sequence 2, Appli
18	7305	58.8	1459	16	US-10-239-498A-13	Sequence 13, Appl
19	7295	58.8	1459	16	US-10-239-498A-15	Sequence 15, Appl
20	7292	58.7	1459	16	US-10-239-498A-4	Sequence 4, Appli
21	7219	58.1	1438	13	US-10-006-091-1	Sequence 1, Appli
22	7219	58.1	1438	13	US-10-047-257-1	Sequence 1, Appli
23	7219	58.1	1438	14	US-10-225-900-1	Sequence 1, Appli
24	6264	50.5	1431	12	US-10-681-970-4	Sequence 4, Appli
25	6264	50.5	1431	13	US-10-095-718-4	Sequence 4, Appli
26	6114	49.2	1443	14	US-10-187-319-39	Sequence 39, Appl
27	6114	49.2	1443	14	US-10-131-510A-39	Sequence 39, Appl
28	2824.5	22.7	2196	14	US-10-115-563-14	Sequence 14, Appl
29	2822.5	22.7	2196	15	US-10-360-101-259	Sequence 259, App
30	2822.5	22.7	2196	14	US-10-172-712-31	Sequence 31, Appl
31	1726	13.9	368	14	US-10-187-319-4	Sequence 4, Appli
32	1726	13.9	368	14	US-10-131-510A-4	Sequence 4, Appli
33	1339.5	10.8	1160	12	US-10-147-493-234	Sequence 234, App
34	1339.5	10.8	1160	12	US-10-145-127-234	Sequence 234, App
35	1339.5	10.8	1160	12	US-10-160-503-234	Sequence 234, App
36	1339.5	10.8	1160	12	US-10-143-118-234	Sequence 234, App
37	1339.5	10.8	1160	12	US-10-144-993-234	Sequence 234, App
38	1339.5	10.8	1160	12	US-10-158-787-234	Sequence 234, App
39	1339.5	10.8	1160	12	US-10-140-024-234	Sequence 234, App
40	1339.5	10.8	1160	14	US-10-028-072-234	Sequence 234, App
41	1339.5	10.8	1160	14	US-10-121-049-234	Sequence 234, App
42	1339.5	10.8	1160	14	US-10-123-904-234	Sequence 234, App
43	1339.5	10.8	1160	14	US-10-140-470-234	Sequence 234, App
44	1339.5	10.8	1160	14	US-10-175-746-234	Sequence 234, App
45	1339.5	10.8	1160	14	US-10-176-918-234	Sequence 234, App

ALIGNMENTS

RESULT 1
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US20030044982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VII:
; FILE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10132,829
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-132-829-4

Query Match	100.0%	Score	12410;	DB	14;	Length	2351;
Best Local Similarity	100.0%	Prod. No.	0;				
Matches	2350;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						
QY	1	MQIELSTCFCLLRFCFSFSA	TRYYLGA	VSWDYMSD	LGELPVD	ARFPPR	PKSPFN 60
Db	1	MQIELSTCFCLLRFCFSFSA	TRYYLGA	VSWDYMSD	LGELPVD	ARFPPR	PKSPFN 60
QY	61	TSVYKTKLTFVEFDHLFNIAKPP	PMGLGPTIO	AEVDTVIT	LKMA	SHFVSLH	AV 120
Db	61	TSVYKTKLTFVEFDHLFNIAKPP	PMGLGPTIO	AEVDTVIT	LKMA	SHFVSLH	AV 120
QY	121	GVSVKASEGAEYDDQTSOREKDD	KVFP	GGSHYV	VOVLKENG	PMASDP	CLTYLSLH 180
Db	121	GVSVKASEGAEYDDQTSOREKDD	KVFP	GGSHYV	VOVLKENG	PMASDP	CLTYLSLH 180

181 QY VDLVKDLSGLIGALLVCREGLAKEKQTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
181 DB VDLVKDLSGLIGALLVCREGLAKEKQTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
241 QY AASARAWPMHTVNGVYVNSRPLGLGCHRSKYVYHVI GMGTTPEVHSIFLEGTPLVRNH 300
241 DB AASARAWPMHTVNGVYVNSRPLGLGCHRSKYVYHVI GMGTTPEVHSIFLEGTPLVRNH 300
301 QY ROASLEIIPITFLTACTLLMDLGOFLSCHEISSHQHDGMEA VVKVDSOPEEPOLRMKNE 360
301 DB ROASLEIIPITFLTACTLLMDLGOFLSCHEISSHQHDGMEA VVKVDSOPEEPOLRMKNE 360
361 QY EAEDYDDDDITDSEMDVVRPDDDNSSPSFTQIRSAVKKHGPKTWVHYIAABEEDWDYAPLVIA 420
361 DB EAEDYDDDDITDSEMDVVRPDDDNSSPSFTQIRSAVKKHGPKTWVHYIAABEEDWDYAPLVIA 420
421 QY PDRSYKSOYLNGQRIQGRKYKVRPMAYTDETFTKTRAIQHESGILGPLLYGEGDTL 480
421 DB PDRSYKSOYLNGQRIQGRKYKVRPMAYTDETFTKTRAIQHESGILGPLLYGEGDTL 480
481 QY LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFPLPGEI FYKXKWTVTVEDGP 540
481 DB LIIFKNOASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPFPLPGEI FYKXKWTVTVEDGP 540
541 QY TKSDPRCLTRYSSVFVNMERDILASGLIGPLLCYKESVDQRGNQIMSKRNVLFSVDFE 600
541 DB TKSDPRCLTRYSSVFVNMERDILASGLIGPLLCYKESVDQRGNQIMSKRNVLFSVDFE 600
601 QY NRSWYLTEINQIFLNFAGVQLEDPEFOASNIMHSINGVYVDSLOLSVCLHEVAVWYILS 660
601 DB NRSWYLTEINQIFLNFAGVQLEDPEFOASNIMHSINGVYVDSLOLSVCLHEVAVWYILS 660
661 QY IGAQTDFLSVFSGVTFKHKWVEDTLTLPFSGETVFMSEMPNGLWILGCHNSDFNRG 720
661 DB IGAQTDFLSVFSGVTFKHKWVEDTLTLPFSGETVFMSEMPNGLWILGCHNSDFNRG 720
721 QY MVALKVSSCDKNTGYVEDSDVEDISAYLLSKNAIEPRSFSONSRHPSTROKQFNATTI 780
721 DB MVALKVSSCDKNTGYVEDSDVEDISAYLLSKNAIEPRSFSONSRHPSTROKQFNATTI 780
781 QY PENDIEKTDPMFAHRTPMKIQNVSSDLMLLRQSPTEHG:SLSDLOBAKYETSDSDPS 840
781 DB PENDIEKTDPMFAHRTPMKIQNVSSDLMLLRQSPTEHG:SLSDLOBAKYETSDSDPS 840
841 QY PGAIDSNNSLSEMTFRPQLHHSQDVFTPESGQLRLNKLGTAA TELKKLDPKVSST 900
841 DB PGAIDSNNSLSEMTFRPQLHHSQDVFTPESGQLRLNKLGTAA TELKKLDPKVSST 900
901 QY SNNLISTISDNLAACTDNITSSLGPPSMVPHYDSQLDITLPGKSSPLTESGGPISLSEB 960
901 DB SNNLISTISDNLAACTDNITSSLGPPSMVPHYDSQLDITLPGKSSPLTESGGPISLSEB 960
961 QY NNDKLLSGLMNSQESSWGKNVSSTESGRLFKGKAHGFALLTKDNALFKYSISLLKTN 1020
961 DB NNDKLLSGLMNSQESSWGKNVSSTESGRLFKGKAHGFALLTKDNALFKYSISLLKTN 1020
1021 QY KTSNNSATNRKTHIDGPSLLIENSPPVWQNILESDETEFKKVTPLIHDRMLMDKNAALRL 1080
1021 DB KTSNNSATNRKTHIDGPSLLIENSPPVWQNILESDETEFKKVTPLIHDRMLMDKNAALRL 1080
1081 QY NMSNKTTSKKNEMVQOKKEGPIPPDAQNPDMSPEKMLFLPESARWIORTHCKNSLNSG 1140
1081 DB NMSNKTTSKKNEMVQOKKEGPIPPDAQNPDMSPEKMLFLPESARWIORTHCKNSLNSG 1140
1141 QY QGSPKQOLVSLGPEKSVGEGQFSLSEKXNVVVGKEFTKDVGLKEMVPPSSRNLFILNLDN 1200
1141 DB QGSPKQOLVSLGPEKSVGEGQFSLSEKXNVVVGKEFTKDVGLKEMVPPSSRNLFILNLDN 1200
1201 QY LHENNTNHOEKKIOEBIEBKETLIOENVLPQ:HTVTGTGNFMKNLFLI:STRQNVEGSYD 1260
1201 DB LHENNTNHOEKKIOEBIEBKETLIOENVLPQ:HTVTGTGNFMKNLFLI:STRQNVEGSYD 1260
1261 QY GAYAPVLODFRSLNDSNRKTKHTAHFSKKEBENLEGLNQTKQIVEKYACTTRISPN 1320

1261 DB GAYAPVLODFRSLNDSNRKTKHTAHFSKKEBENLEGLNQTKQIVEKYACTTRISPN 1320
1321 QY SQQNFVTORSKRALQORLPLEETELBKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
1321 DB SQQNFVTORSKRALQORLPLEETELBKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
1381 QY KGAITQSPLSCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTRVLFPQDNSSHLPAASY 1440
1381 DB KGAITQSPLSCLTRSHSIPQANRSPPIAKVSSPPSIRPIYLTRVLFPQDNSSHLPAASY 1440
1441 QY RKXGSGVOESSHPIQGAHKONLSAILTLEMTGQOREVGSIGTSATNSVTYKKVENTVLP 1500
1441 DB RKXGSGVOESSHPIQGAHKONLSAILTLEMTGQOREVGSIGTSATNSVTYKKVENTVLP 1500
1501 QY K3DLPKXTSGKVELLPKVHIYQKDLFPPTSTNSGPGHLDVEGSLLOQTGEGAIKWNEANRP 1560
1501 DB K3DLPKXTSGKVELLPKVHIYQKDLFPPTSTNSGPGHLDVEGSLLOQTGEGAIKWNEANRP 1560
1561 QY GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKQESKPEKTAFFKXOTILSL 1620
1561 DB GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKQESKPEKTAFFKXOTILSL 1620
1621 QY NACESNHAIAAINEGONKPEIEVWAKQGRTERLCSQNPVILKHHORBITRTTLOSDOEB 1680
1621 DB NACESNHAIAAINEGONKPEIEVWAKQGRTERLCSQNPVILKHHORBITRTTLOSDOEB 1680
1681 QY DYDDTISVENKKEFDIYDEDENOSPRSFQKTRHYFIAAVERLDYCMSSSPHVLNRR 1740
1681 DB DYDDTISVENKKEFDIYDEDENOSPRSFQKTRHYFIAAVERLDYCMSSSPHVLNRR 1740
1741 QY AQSGSVPOFKXVVOFETDGSFTQPLYRGELNEHGLLGPYIRAEVEDNIMVTFNQNAR 1800
1741 DB AQSGSVPOFKXVVOFETDGSFTQPLYRGELNEHGLLGPYIRAEVEDNIMVTFNQNAR 1800
1801 QY PYSFYSSLIISVEEDORQGAERPKNFVKNEKTYEYKQHHMAPTKDPDCKAWAYFSDV 1860
1801 DB PYSFYSSLIISVEEDORQGAERPKNFVKNEKTYEYKQHHMAPTKDPDCKAWAYFSDV 1860
1861 QY DIEKDVHSLGILGPLLVCHTNTLNPAGHQVTVQEPALFPTTDETKSWYFTENMERNCR 1920
1861 DB DIEKDVHSLGILGPLLVCHTNTLNPAGHQVTVQEPALFPTTDETKSWYFTENMERNCR 1920
1921 QY PCNIQMEDPTFKENTRFAALNGYIMDTLPGVWAOQRIWYLLSMGNENIHSIHFGH 1980
1921 DB PCNIQMEDPTFKENTRFAALNGYIMDTLPGVWAOQRIWYLLSMGNENIHSIHFGH 1980
1981 QY VFTVRKKEEYKVALYNLYPGVETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
1981 DB VFTVRKKEEYKVALYNLYPGVETVEMLPKAGIWRVECLIGEHLHAGMSTFLVYSNKC 2040
2041 QY QTPLGWASGHIRDPQITASGGYQGWAPKLARLHYSGSINAWSTKEFFSWIKVDLLAPMII 2100
2041 DB QTPLGWASGHIRDPQITASGGYQGWAPKLARLHYSGSINAWSTKEFFSWIKVDLLAPMII 2100
2101 QY HGKIQGAROKSSLIYSQPIIMYSLDGKKWQTYGNSTGLMWFGNVDSGSIKHNIEN 2160
2101 DB HGKIQGAROKSSLIYSQPIIMYSLDGKKWQTYGNSTGLMWFGNVDSGSIKHNIEN 2160
2161 QY PPIIARYIRLHPHTHYSIRSTRMELMGCDLNSCPLGMESKAI:SDAQITASSYTNMFA 2220
2161 DB PPIIARYIRLHPHTHYSIRSTRMELMGCDLNSCPLGMESKAI:SDAQITASSYTNMFA 2220
2221 QY TWSPSKARLHLQGRSNARPOVNNPKWLQVDFOKTMKVTVGTTQGVKSLITSMYVKRPL 2280
2221 DB TWSPSKARLHLQGRSNARPOVNNPKWLQVDFOKTMKVTVGTTQGVKSLITSMYVKRPL 2280
2281 QY ISSSQDGHQWTLFPQNGKVKVQGNQDSFTPVNSLDPPLLTLYLRIHPQSWHQAIALRM 2340
2281 DB ISSSQDGHQWTLFPQNGKVKVQGNQDSFTPVNSLDPPLLTLYLRIHPQSWHQAIALRM 2340
2341 QY EYLGCEAODLY 2351

Db 2341 EVLGEAQDLY 2351

RESULT 2

US-10-172-712-27

; Sequence 27, Application US/10172712

; Publication No. US20030125232A1

; GENERAL INFORMATION:

; APPLICANT: GRIFPIN, JOHN H.

; APPLICANT: GALE, ANDREW J.

; APPLICANT: GETZORE, ELIZABETH D.

; APPLICANT: PELLEQUER, JEAN-LUC

; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS

; FILE REFERENCE: 4198-4001US1

; CURRENT APPLICATION NUMBER: US/10/172,712

; CURRENT FILING DATE: 2002-09-30

; PRIOR APPLICATION NUMBER: 60/298,578

; PRIOR FILING DATE: 2001-06-14

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 27

; LENGTH: 2351

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-172-712-27

Query Match 100.0%; Score 12410; DB 14; Length 2351;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MQELSTCFCLLRPCFSATRRYYLGAVELSWDYNQSDLGELPVDARPPRPVKSPFN 60

Db 1 MQELSTCFCLLRPCFSATRRYYLGAVELSWDYNQSDLGELPVDARPPRPVKSPFN 60

QY 61 TSVVYKTKLFVEFTDHLNIAKPRPPWMLGCLPTTQAEVVDVTVITLKNMASHPVSLHAV 120

Db 61 TSVVYKTKLFVEFTDHLNIAKPRPPWMLGCLPTTQAEVVDVTVITLKNMASHPVSLHAV 120

QY 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGGSHTYVWQVLKENGPMASDPLCLTYSLSH 180

Db 121 GVSYWKASEGAEYDDQTSQREKDDKVPFGGSHTYVWQVLKENGPMASDPLCLTYSLSH 180

QY 181 VDLVKDLSNGIIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240

Db 181 VDLVKDLSNGIIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240

QY 241 AASAPAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGWGTTPVHSTFLECHTFLVRNH 300

Db 241 AASAPAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGWGTTPVHSTFLECHTFLVRNH 300

QY 301 ROASLEISPIITFLTAQTLMDLQGLLSCHISSHQHGDGEAYVKVDSCPEBPQLRMKNE 360

Db 301 ROASLEISPIITFLTAQTLMDLQGLLSCHISSHQHGDGEAYVKVDSCPEBPQLRMKNE 360

QY 361 EADYDDDLTISEMDVVPDDNDSPTQIRSAVKHPTKTHWYTAASEBWDVAPLVIA 420

Db 361 EADYDDDLTISEMDVVPDDNDSPTQIRSAVKHPTKTHWYTAASEBWDVAPLVIA 420

QY 421 PDRSYKSOYLNGPQRTGRKYKVRWAYTDETFKTREAIQHESGILGPLLGEVGDTL 480

Db 421 PDRSYKSOYLNGPQRTGRKYKVRWAYTDETFKTREAIQHESGILGPLLGEVGDTL 480

QY 481 LILFKNQASRPNTYPHGIDTVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVVEDGP 540

Db 481 LILFKNQASRPNTYPHGIDTVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVVEDGP 540

QY 541 TKSDBRCUTRYSSFWNMRDIALSGILGELLICYKESVDQRGNQMSDKRNVLFSVDE 600

Db 541 TKSDBRCUTRYSSFWNMRDIALSGILGELLICYKESVDQRGNQMSDKRNVLFSVDE 600

QY 601 NRSWLTENIQRFLEPNPAGVQLEDPEFQASNMHNSINGVVPDLSQLSVCLHEVAYWYILS 660

Db 601 NRSWLTENIQRFLEPNPAGVQLEDPEFQASNMHNSINGVVPDLSQLSVCLHEVAYWYILS 660

QY 661 IGAOTDFLSVFFSCYTFFKHVMVYEDTTLTPFSGETVFMSENPGLMILGCHNSDFNRG 720

Db 661 IGAOTDFLSVFFSCYTFFKHVMVYEDTTLTPFSGETVFMSENPGLMILGCHNSDFNRG 720

QY 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI 780

Db 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI 780

QY 781 PENDIEKTDPMFAHRTMPKIQNVSSDGLMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840

Db 781 PENDIEKTDPMFAHRTMPKIQNVSSDGLMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840

QY 841 PGALDSNNLSSEMTHFRPQLHHSQDMVFTPESGQLRNLNEKLGTTAATELKKLDFKVSST 900

Db 841 PGALDSNNLSSEMTHFRPQLHHSQDMVFTPESGQLRNLNEKLGTTAATELKKLDFKVSST 900

QY 901 SNNLIISTPSDNLAAGTDNTSSLGPPSPVHYDSQLDITTLFGKXSSPLTESGGPLSSEE 960

Db 901 SNNLIISTPSDNLAAGTDNTSSLGPPSPVHYDSQLDITTLFGKXSSPLTESGGPLSSEE 960

QY 961 NNDKLLIESGLMNSOESSWGKQVSSSTESGRLPKGRAGPALLTKDNALFKVSLSLKTN 1020

Db 961 NNDKLLIESGLMNSOESSWGKQVSSSTESGRLPKGRAGPALLTKDNALFKVSLSLKTN 1020

QY 1021 KTSNNSATNRKTHIDGFSLLIENSVPWQNILEDSTEFKVTPLIHDMMLMDKNATALRL 1080

Db 1021 KTSNNSATNRKTHIDGFSLLIENSVPWQNILEDSTEFKVTPLIHDMMLMDKNATALRL 1080

QY 1081 NHMKNKTTSSKNMVMQCKKEGPIPPDAQNPDMSFFKMLFLPESARMIQRTGHKNSLNSG 1140

Db 1081 NHMKNKTTSSKNMVMQCKKEGPIPPDAQNPDMSFFKMLFLPESARMIQRTGHKNSLNSG 1140

QY 1141 QGPSPKOLVSLGPEKSVGQNFSEKKNVVGKEFTKQVCLKEMVFPSSNLFITWLDN 1200

Db 1141 QGPSPKOLVSLGPEKSVGQNFSEKKNVVGKEFTKQVCLKEMVFPSSNLFITWLDN 1200

QY 1201 LHENNTNMQEKKIOBEIEKKTLLIQENVVLPOIHVTGTGNFMKNLFLLSSTRQNVESYD 1260

Db 1201 LHENNTNMQEKKIOBEIEKKTLLIQENVVLPOIHVTGTGNFMKNLFLLSSTRQNVESYD 1260

QY 1261 GAYAPVLQDFPSLNDSTNRKTAHPSKKEENLEGLNQOTQOIVEKYACTTRISNT 1320

Db 1261 GAYAPVLQDFPSLNDSTNRKTAHPSKKEENLEGLNQOTQOIVEKYACTTRISNT 1320

QY 1321 SQQNFVQTSKRALQKQRLPEETELEKRIIVDDTSTQWSKNMKHLPTSTLTQIDYNEKE 1380

Db 1321 SQQNFVQTSKRALQKQRLPEETELEKRIIVDDTSTQWSKNMKHLPTSTLTQIDYNEKE 1380

QY 1381 KGAITQSPSLDCLTRSHSIPOANRSPPLIAKVSSFPPIRPIYLTIRVLFDQNSSHLPAASY 1440

Db 1381 KGAITQSPSLDCLTRSHSIPOANRSPPLIAKVSSFPPIRPIYLTIRVLFDQNSSHLPAASY 1440

QY 1441 RKDGSVQESHSFLOGAKKNLSAILTLEMTGQREVSGLSGTATNSVTYKKVENVLP 1500

Db 1441 RKDGSVQESHSFLOGAKKNLSAILTLEMTGQREVSGLSGTATNSVTYKKVENVLP 1500

QY 1501 KPDLPKTSKVELLPKVHIYOKDLFPTETSNPSGHLDLVEGSLLOQTEGAIKWEANRP 1560

Db 1501 KPDLPKTSKVELLPKVHIYOKDLFPTETSNPSGHLDLVEGSLLOQTEGAIKWEANRP 1560

QY 1561 GKVPFLVATESAKTSPKLLDPLAWNHYGTQIPKEEWKSOEKSPEKTAFFKKDITLSL 1620

Db 1561 GKVPFLVATESAKTSPKLLDPLAWNHYGTQIPKEEWKSOEKSPEKTAFFKKDITLSL 1620

QY 1621 NACSNHAIAINBQONKPEIEVTWAKQRTERLCSONPVLKHOBEIRITTTIQSQEE 1680

Db 1621 NACSNHAIAINBQONKPEIEVTWAKQRTERLCSONPVLKHOBEIRITTTIQSQEE 1680

QY 1681 IDYDOTTIVEMKKEDFDIYDEENQSPRSPOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740

Db 1681 IDYDOTTIVEMKKEDFDIYDEENQSPRSPOKTRHYFIAAVERLWDYGMSSSPHVLNR 1740

1741 AQSGVPOFKKVVFOEFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFNQASR 1800
Db
1741 AQSGVPOFKKVVFOEFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFNQASR 1800
Qy
1801 PYSFYSSLSYEDORQAGAPKRNFKVKNETKTYFKVQHNAPTKDBPDCKAWAYFSDV 1860
Db
1801 PYSFYSSLSYEDORQAGAPKRNFKVKNETKTYFKVQHNAPTKDBPDCKAWAYFSDV 1860
Qy
1861 DLEKDVHSLGELGLVCHNTNLPAGHROVTVQESALPFTIFDETKSWYFTENMERNCRA 1920
Db
1861 DLEKDVHSLGELGLVCHNTNLPAGHROVTVQESALPFTIFDETKSWYFTENMERNCRA 1920
Qy
1921 PCNIQWEDPTFKENYFAHNGYIMDTPLGLVMAQDQRIWYLLSMGNSNENHSIHFSGH 1980
Db
1921 PCNIQWEDPTFKENYFAHNGYIMDTPLGLVMAQDQRIWYLLSMGNSNENHSIHFSGH 1980
Qy
1981 VFTVRKKEBYKVALYNLYPGVETVEMLPKAGIWEVECLI GEHLHAGNSTFLVYSNKC 2040
Db
1981 VFTVRKKEBYKVALYNLYPGVETVEMLPKAGIWEVECLI GEHLHAGNSTFLVYSNKC 2040
Qy
2041 QTPGLWASGHIRDFOITASGOYGOWAPKLARLHYSOSINAWSTKBPFSWIKVDLLAPMII 2100
Db
2041 QTPGLWASGHIRDFOITASGOYGOWAPKLARLHYSOSINAWSTKBPFSWIKVDLLAPMII 2100
Qy
2101 HGKIQGARQKSSLYISOFLIIMYSLDGKKWQTYRGNSTGTLMVPPGVNDSSGIKENIFN 2160
Db
2101 HGKIQGARQKSSLYISOFLIIMYSLDGKKWQTYRGNSTGTLMVPPGVNDSSGIKENIFN 2160
Qy
2161 PPIIARYIRLPHPHYSIRSTLRMELMGCDLNSCMLPGWESKAISDAQITASSYFTNMPA 2220
Db
2161 PPIIARYIRLPHPHYSIRSTLRMELMGCDLNSCMLPGWESKAISDAQITASSYFTNMPA 2220
Qy
2221 TWSPKARLHLQGRSNAMPPOVNNPKWLQVDFQKTKMKTGVTITQGVKSLLTSMYVKEPL 2280
Db
2221 TWSPKARLHLQGRSNAMPPOVNNPKWLQVDFQKTKMKTGVTITQGVKSLLTSMYVKEPL 2280
Qy
2281 ISSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDPPLTRYLRPHQSWVHQIALRM 2340
Db
2281 ISSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDPPLTRYLRPHQSWVHQIALRM 2340
Qy
2341 EVLGCEAQDLY 2351
Db
2341 EVLGCEAQDLY 2351
RESULT 3
US-10-133-907-4
; Sequence 4, Application US/10133907
; Publication No. US2003019523A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-Pal170
; CURRENT APPLICATION NUMBER: US/10/133,907
; CURRENT FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: 60/286,314
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-133-907-4
Query Match 100.0%; Score 12410; DB 14; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYNQSDLGELPVDARPPRPVKSPFFN 60

1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYNQSDLGELPVDARPPRPVKSPFFN 60
Qy
61 TSVVYKXTLFVSEFTDHLNIAKPRPPWMGLLGPITQAEVYDVTVTITLKNMASHPVSLHAV 120
Db
61 TSVVYKXTLFVSEFTDHLNIAKPRPPWMGLLGPITQAEVYDVTVTITLKNMASHPVSLHAV 120
Qy
121 GYSYWKASGAEAYDDQTSQREKEDDKVPFGSGSHYTVQVLKENGPMASDPCLCLTYSLYLSH 180
Db
121 GYSYWKASGAEAYDDQTSQREKEDDKVPFGSGSHYTVQVLKENGPMASDPCLCLTYSLYLSH 180
Qy
181 VDLVXDLNSGLIGALLVOREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
Db
181 VDLVXDLNSGLIGALLVOREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
Qy
241 AASARAWPMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLECHTEFLVRNH 300
Db
241 AASARAWPMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLECHTEFLVRNH 300
Qy
301 ROASLEISPIFLTAQTLMDLIGQFLLSCHISSHQHDGMEAYVKYVDSCEPBPQLRMKNE 360
Db
301 ROASLEISPIFLTAQTLMDLIGQFLLSCHISSHQHDGMEAYVKYVDSCEPBPQLRMKNE 360
Qy
361 EABDYDDDLTDSMDVVRFDNDSPTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
Db
361 EABDYDDDLTDSMDVVRFDNDSPTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
Qy
421 PDRSYKQYLNNGPQRIGRYKVKVRFNAYTDETPKTEAZQHESGILGLPLYGEGVDTL 480
Db
421 PDRSYKQYLNNGPQRIGRYKVKVRFNAYTDETPKTEAZQHESGILGLPLYGEGVDTL 480
Qy
481 LLIIFXQASRPNIYPHGITDVRPIYSRRLPKGVKHLKOPFIPGLPEIKYKWTVTVEGSP 540
Db
481 LLIIFXQASRPNIYPHGITDVRPIYSRRLPKGVKHLKOPFIPGLPEIKYKWTVTVEGSP 540
Qy
541 TKSDPRCLTRYSSVFNVERDLASGLIGPLLCYKESVDQGNQIMSKNVILFSVDE 600
Db
541 TKSDPRCLTRYSSVFNVERDLASGLIGPLLCYKESVDQGNQIMSKNVILFSVDE 600
Qy
601 NRSWYLTENIQRFLENPAGVQLEDEPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYLS 660
Db
601 NRSWYLTENIQRFLENPAGVQLEDEPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYLS 660
Qy
661 IGAQTDPLSVFPFGYTFKHQWYEDTLTLPFSGETVFMSENPGIWLICHNSDFNRG 720
Db
661 IGAQTDPLSVFPFGYTFKHQWYEDTLTLPFSGETVFMSENPGIWLICHNSDFNRG 720
Qy
721 MTALLKVSQCDKNTGDYVEDSVEDISAYLLSKNNAIEPRSPSONSRHSTROKQFNATTI 780
Db
721 MTALLKVSQCDKNTGDYVEDSVEDISAYLLSKNNAIEPRSPSONSRHSTROKQFNATTI 780
Qy
781 PENDIEKTDPPFAHRTPMFKIQNVSSSLLMLLRQSPTHGLSLSDLOEAKYETSDSPS 840
Db
781 PENDIEKTDPPFAHRTPMFKIQNVSSSLLMLLRQSPTHGLSLSDLOEAKYETSDSPS 840
Qy
841 PGAISSNLSLSEMTFRPQLHSHSGDMVTPESGLQRLNEKLGTTAAATLKLDPKVSST 900
Db
841 PGAISSNLSLSEMTFRPQLHSHSGDMVTPESGLQRLNEKLGTTAAATLKLDPKVSST 900
Qy
901 SNLLISTIPSDNLAAGTNTSSLGPPSPVHYDSQDITTLFGKSSPLTETSGGPLSLSEE 960
Db
901 SNLLISTIPSDNLAAGTNTSSLGPPSPVHYDSQDITTLFGKSSPLTETSGGPLSLSEE 960
Qy
961 NNDSKLLSGLMNSQESSWGKRVSTESGRIFKGRAGHPALLTKDNALFKVSIILLKTN 1020
Db
961 NNDSKLLSGLMNSQESSWGKRVSTESGRIFKGRAGHPALLTKDNALFKVSIILLKTN 1020
Qy
1021 KTSNKSATNRKTHIDGPSLITENSVMQNTLESDETEKVTPLIHDMLMDKNTAURL 1080
Db
1021 KTSNKSATNRKTHIDGPSLITENSVMQNTLESDETEKVTPLIHDMLMDKNTAURL 1080
Qy
1081 NWSNKTTSKKNMVMQKKEGPIPPDAQNPVSPFFKMLFPESARWQRTHGKXSLNSG 1140
Db
1081 NWSNKTTSKKNMVMQKKEGPIPPDAQNPVSPFFKMLFPESARWQRTHGKXSLNSG 1140

1141 QGSPKQVLSIGPEKVEGQNFSEKKNVVGKGEFTKDVGLKEMVPPSSRNLFNTLNL 1200
1141 QGSPKQVLSIGPEKVEGQNFSEKKNVVGKGEFTKDVGLKEMVPPSSRNLFNTLNL 1200
1201 LHENNTNHEKKTQIEIEKKEKTIQENVVLPOIHTVTGKTNMKNLFLLSRQNVGSYD 1260
1201 LHENNTNHEKKTQIEIEKKEKTIQENVVLPOIHTVTGKTNMKNLFLLSRQNVGSYD 1260
1261 GAYAPVLOPFRSLNDSNTKHTAHFSKKGEBENLEGLGNQTKQIVKVIACITRISPT 1320
1261 GAYAPVLOPFRSLNDSNTKHTAHFSKKGEBENLEGLGNQTKQIVKVIACITRISPT 1320
1321 SOONFVTSKQKALKOPRLPLEBTELEKRIIVDDTSTOWSKMKHLTPSTLTQIDYNEK 1380
1321 SOONFVTSKQKALKOPRLPLEBTELEKRIIVDDTSTOWSKMKHLTPSTLTQIDYNEK 1380
1381 KGATQSPISDCLTRSHSIPQANRSPPIAKVSSPFSIRPIYLVTRVLFQDNSSHLPAAS 1440
1381 KGATQSPISDCLTRSHSIPQANRSPPIAKVSSPFSIRPIYLVTRVLFQDNSSHLPAAS 1440
1441 RKKDSGVQSSSHFLOCAKKNLSALTLEMTGDQREVGLSATNSVTYKKVENTVLP 1500
1441 RKKDSGVQSSSHFLOCAKKNLSALTLEMTGDQREVGLSATNSVTYKKVENTVLP 1500
1501 KEDLPKTSQKVELLPKVHYQKDLFPETSNQSPGHLDLVEGSLQCTEGALKWNEANRP 1560
1501 KEDLPKTSQKVELLPKVHYQKDLFPETSNQSPGHLDLVEGSLQCTEGALKWNEANRP 1560
1561 GXVPFLRVATSAKTPSKLLDPLANDNHYGTQIPKEWKSEKSPKTAFFKKOTILSL 1620
1561 GXVPFLRVATSAKTPSKLLDPLANDNHYGTQIPKEWKSEKSPKTAFFKKOTILSL 1620
1621 NACESHAAIAANEGONKPEIETWAKQGRTELCSQNPVLKSHORELTRTLOSDDSE 1680
1621 NACESHAAIAANEGONKPEIETWAKQGRTELCSQNPVLKSHORELTRTLOSDDSE 1680
1681 IDYDDTISVEMKEDFDIYDEDENSQSPRQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1681 IDYDDTISVEMKEDFDIYDEDENSQSPRQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
1741 AQSGVQPKVQKVVQFTDGSSTQPIYRGELNEHLLGPGYTRAEVEDNIMVTFRQASR 1800
1741 AQSGVQPKVQKVVQFTDGSSTQPIYRGELNEHLLGPGYTRAEVEDNIMVTFRQASR 1800
1801 PYSFYSSLISYEDDQQAEPKRNFKVNETKTYFMKVQHVMAPTKDFDCKAWAFSDV 1860
1801 PYSFYSSLISYEDDQQAEPKRNFKVNETKTYFMKVQHVMAPTKDFDCKAWAFSDV 1860
1861 DLEKDVHSLIGPLLVCHTNTLNPAGHROVTVQEFALFTTIDETKSWYFTENMERNCA 1920
1861 DLEKDVHSLIGPLLVCHTNTLNPAGHROVTVQEFALFTTIDETKSWYFTENMERNCA 1920
1921 PCNIQWEDTFKENTRPHANGYIMDTLGLVMAQDORRWLLSGNSNETHSHFSQH 1980
1921 PCNIQWEDTFKENTRPHANGYIMDTLGLVMAQDORRWLLSGNSNETHSHFSQH 1980
1981 VFTVRKKEEYKMAALNLYPGVETVEMLPKAGIWRVECLIGHLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKMAALNLYPGVETVEMLPKAGIWRVECLIGHLHAGMSTLFLVYSNKC 2040
2041 QTPLGWAGHIRDFOITAGSQGWAPKRLARLHYSGINAWSTKEPFSWKVDLAPMII 2100
2041 QTPLGWAGHIRDFOITAGSQGWAPKRLARLHYSGINAWSTKEPFSWKVDLAPMII 2100
2101 HGKTCGAKQKSSLYISQFIIMSLDGKKWQTYRGNSTGTLWFFGNVDSSGIKINIEN 2160
2101 HGKTCGAKQKSSLYISQFIIMSLDGKKWQTYRGNSTGTLWFFGNVDSSGIKINIEN 2160
2161 PPIIARYIELPCHYISIRGTRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMPA 2220
2161 PPIIARYIELPCHYISIRGTRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMPA 2220

RESULT 4

US-10-411-037-30
; Sequence 30, Application US/10411037
; Publication NO. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Bowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; TYPE: PRT
; LENGTH: 2351
; ORGANISM: Homo sapiens
US-10-411-037-30

Query Match 99.9%; Score 12401; DB 12; Length 2351;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MOELSTCFFCLLAFRCFSATRRYYLGAVALSWDQSDIGELPVDARPPBPVPSFPPN 60
DB 1 MOELSTCFFCLLAFRCFSATRRYYLGAVALSWDQSDIGELPVDARPPBPVPSFPPN 60
QY 61 TSVVYKTLFVEFTDHLFNIAPRPPMGLLGPTTQAEVYDTVTWTLKNMASHPVSLHAV 120
DB 61 TSVVYKTLFVEFTDHLFNIAPRPPMGLLGPTTQAEVYDTVTWTLKNMASHPVSLHAV 120
QY 121 GVSYWKASEGAEYDDQTSOREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYL 180
DB 121 GVSYWKASEGAEYDDQTSOREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYL 180
QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPRMGTWNGYVNRSLPGLICHRKSVYWHVIGXGTTPEVHSIFLEGHTFLVRNH 300

Db 241 AASARAPMHTVNGVVRSLPCLGCRKSKYVWVINGCTPVSHSIFLEHTELVNH 300
Qy 301 ROASLEISPIITELTAQTLMDLQFLLSCHSHOHGMEAVVYKVDSPCEBPOLRMKNE 360
Db 301 ROASLEISPIITELTAQTLMDLQFLLSCHSHOHGMEAVVYKVDSPCEBPOLRMKNE 360
Qy 361 BAEDYDDDLTDSMDVAFDDNSPSFQIRSAVAKKPKTWVHYIAAEEEDMDYAPLVLA 420
Db 361 BAEDYDDDLTDSMDVAFDDNSPSFQIRSAVAKKPKTWVHYIAAEEEDMDYAPLVLA 420
Qy 421 PDORSVKSOYLANGPORGRKYKVRFMAYTDETFKTRAIQHESGILGPLLYGEGVDTL 480
Db 421 PDORSVKSOYLANGPORGRKYKVRFMAYTDETFKTRAIQHESGILGPLLYGEGVDTL 480
Qy 481 LIIFXNQASRPNIYPHGTDVVRPYSRRLPKGVKHLKDPFPLPGEIFKYKVTVVEDGP 540
Db 481 LIIFXNQASRPNIYPHGTDVVRPYSRRLPKGVKHLKDPFPLPGEIFKYKVTVVEDGP 540
Qy 541 TKSDPRECJTRYSSFVNMERDLASGLIGPLLYCYKESVDORGNOJMSDKENILSVFDE 600
Db 541 TKSDPRECJTRYSSFVNMERDLASGLIGPLLYCYKESVDORGNOJMSDKENILSVFDE 600
Qy 601 NRSWILTENIQRLPNPAGVQLEDPEFQASNIWHSINGVVFDSLQLSVCLHEVAYWYLLS 660
Db 601 NRSWILTENIQRLPNPAGVQLEDPEFQASNIWHSINGVVFDSLQLSVCLHEVAYWYLLS 660
Qy 661 IGAQTDPLSVFSGYTFKHKMYVEDTLTLPFSGETVFNMSMENPGLWILGCHNSDFNRG 720
Db 661 IGAQTDPLSVFSGYTFKHKMYVEDTLTLPFSGETVFNMSMENPGLWILGCHNSDFNRG 720
Qy 721 MTALLKVSSCDKNTGVDYEDSDISAYLLSKNAIEPSSFSONSRHPTROKQFNATTI 780
Db 721 MTALLKVSSCDKNTGVDYEDSDISAYLLSKNAIEPSSFSONSRHPTROKQFNATTI 780
Qy 781 PENDIEKTDPFARHTPMPIKQNVSSDDLMLLRQSPTPHGLSLDLOBAKYETPSDDPS 840
Db 781 PENDIEKTDPFARHTPMPIKQNVSSDDLMLLRQSPTPHGLSLDLOBAKYETPSDDPS 840
Qy 841 PGALDSNNLSLSETHFRPQLHHSQDMVFTPESGLOJRLNEKLGTTAATBLKLDKFSVST 900
Db 841 PGALDSNNLSLSETHFRPQLHHSQDMVFTPESGLOJRLNEKLGTTAATBLKLDKFSVST 900
Qy 901 SNNLSTIPSDNLAACTDNTSSIGPPSPVHYDSQDITLFGKKSPLTESGSPLSLSE 960
Db 901 SNNLSTIPSDNLAACTDNTSSIGPPSPVHYDSQDITLFGKKSPLTESGSPLSLSE 960
Qy 961 NNDKLLSGLMNSQESSWGKNVSTESGRLFKGZAHGPALLTKDNALFKVSI SLLKTN 1020
Db 961 NNDKLLSGLMNSQESSWGKNVSTESGRLFKGZAHGPALLTKDNALFKVSI SLLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGFSLLIENSPPVWONILSDTEPKVTPLIHDMMDKNATALRL 1080
Db 1021 KTSNNSATNRKTHIDGFSLLIENSPPVWONILSDTEPKVTPLIHDMMDKNATALRL 1080
Qy 1081 NMSNKTSSKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPSARWIQRTHKNSLNSG 1140
Db 1081 NMSNKTSSKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPSARWIQRTHKNSLNSG 1140
Qy 1141 QGSPKQLSLGPESKVEGONFISEKNKVVGKGETKOVGLKENYPPSSRNILFLNLON 1200
Db 1141 QGSPKQLSLGPESKVEGONFISEKNKVVGKGETKOVGLKENYPPSSRNILFLNLON 1200
Qy 1201 LHENNTNQEKKIQEELEKXETILIQENVLPQHTVTGKNFMKNLFLLSSTRONVSGSYD 1260
Db 1201 LHENNTNQEKKIQEELEKXETILIQENVLPQHTVTGKNFMKNLFLLSSTRONVSGSYD 1260
Qy 1261 GAYAPVLQDPSRLNDSTNRKTHTAHFSKKGEBENLEGLNQTKQIVEKYACTRISPNT 1320
Db 1261 GAYAPVLQDPSRLNDSTNRKTHTAHFSKKGEBENLEGLNQTKQIVEKYACTRISPNT 1320
Qy 1321 SQNFVTVQSRKALKQFRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

Db 1321 SQNFVTVQSRKALKQFRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGATQSPLSCLTRSHSIPQANSPPIAKVSSPPSRPIYLTRVLFPQDNSSHLPAASY 1440
Db 1381 KGATQSPLSCLTRSHSIPQANSPPIAKVSSPPSRPIYLTRVLFPQDNSSHLPAASY 1440
Qy 1441 RKDSGVQESSHFFQGAKNKNSLAITLLEMTGQOREVSGSLGTSATNSVTYKVENTVLP 1500
Db 1441 RKDSGVQESSHFFQGAKNKNSLAITLLEMTGQOREVSGSLGTSATNSVTYKVENTVLP 1500
Qy 1501 KPDLPKTSGKVELLPKVHIYQKDLFPPTTSGSPGHLDLVEGSLLOQTEGAIKWEANRP 1560
Db 1501 KPDLPKTSGKVELLPKVHIYQKDLFPPTTSGSPGHLDLVEGSLLOQTEGAIKWEANRP 1560
Qy 1561 GKVPFLRVATESSAKTPSKLADPLAWDNHYGTQPKBEWKSQESKPEKTAFKKDTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLADPLAWDNHYGTQPKBEWKSQESKPEKTAFKKDTILSL 1620
Qy 1621 NACESNHAIAAINEGONKPEIEVTWAKQGRTERLCSQNPPLVKHOREITITLQSDQEE 1680
Db 1621 NACESNHAIAAINEGONKPEIEVTWAKQGRTERLCSQNPPLVKHOREITITLQSDQEE 1680
Qy 1681 IDYDITI SVENKKSDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLRNR 1740
Db 1681 IDYDITI SVENKKSDFDIYDEENQSPRSFQKTRHYFIAAVERLWDYGMSSPHVLRNR 1740
Qy 1741 AQSGSVPOFKVVFQFETDGSFTQPLVRGELNEHLLGLGPVIRAEVEDNIWVTFNQASR 1800
Db 1741 AQSGSVPOFKVVFQFETDGSFTQPLVRGELNEHLLGLGPVIRAEVEDNIWVTFNQASR 1800
Qy 1801 PYSFYSSLI SYEEDQORQAEPRKPNFKVKNETKTYFWQVHMAPTKDFDCWAYVSDV 1860
Db 1801 PYSFYSSLI SYEEDQORQAEPRKPNFKVKNETKTYFWQVHMAPTKDFDCWAYVSDV 1860
Qy 1861 DLEKDVHGLIGPLLVCHTNILNPAHGQVTVQEFALFETIPDETCKSHYFENMERNCEA 1920
Db 1861 DLEKDVHGLIGPLLVCHTNILNPAHGQVTVQEFALFETIPDETCKSHYFENMERNCEA 1920
Qy 1921 PCNIQMEPTFKENYRFAINGYIMDTLPLVMAOQRIWYLLSGMGNENIHSIHFSGH 1980
Db 1921 PCNIQMEPTFKENYRFAINGYIMDTLPLVMAOQRIWYLLSGMGNENIHSIHFSGH 1980
Qy 1981 VFTVRKKEBYKMALYNLPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEBYKMALYNLPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Qy 2041 QTPPLGWSGHIRDPQITASQYQWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPPLGWSGHIRDPQITASQYQWAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMII 2100
Qy 2101 HGKIQGARQKFSLSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGSIKHNFN 2160
Db 2101 HGKIQGARQKFSLSYISOFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSGSIKHNFN 2160
Qy 2161 PPIIARYIRLHPTHYSIRSTRMELMGCDLNSCSPLGMESKAI SDAQITASSYTNMFA 2220
Db 2161 PPIIARYIRLHPTHYSIRSTRMELMGCDLNSCSPLGMESKAI SDAQITASSYTNMFA 2220
Qy 2221 TWSPSKARLHLOGRSNAPQVNNPEKMOYDFOKTMKVTVTTQGVKSLTSMVYKBPFL 2280
Db 2221 TWSPSKARLHLOGRSNAPQVNNPEKMOYDFOKTMKVTVTTQGVKSLTSMVYKBPFL 2280
Qy 2281 ISSSQDGHQWTLFFQNGKVKVFCQGNQDFTPVVNSLDPLLTRYLRIHPQSVHQAIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFCQGNQDFTPVVNSLDPLLTRYLRIHPQSVHQAIALRM 2340
Qy 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

	; Sequence 30, Application US/10411026	
	; Publication No. US2004006391A1	
	; GENERAL INFORMATION:	
	; APPLICANT: Neose Technologies, Inc.	
	; INVENTOR: DeFreese, Shawn	
	; APPLICANT: Zopf, David	
	; APPLICANT: Bayer, Robert	
	; APPLICANT: Hakes, David	
	; APPLICANT: Chen, Xi	
	; TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE	
	; FILE REFERENCE: 040853-01-5053	
	; CURRENT APPLICATION NUMBER: US/10/411,026	
	; PRIOR FILING DATE: 2003-04-09	
	; PRICE APPLICATION NUMBER: US 60/328,523	
	; PRIOR FILING DATE: 2001-10-10	
	; PRIOR APPLICATION NUMBER: US 60/344,692	
	; PRIOR FILING DATE: 2001-10-19	
	; PRIOR APPLICATION NUMBER: US 60/387,292	
	; PRIOR FILING DATE: 2002-06-07	
	; PRIOR APPLICATION NUMBER: US 60/391,777	
	; PRIOR FILING DATE: 2002-06-25	
	; PRIOR APPLICATION NUMBER: US 60/396,594	
	; PRIOR FILING DATE: 2002-07-17	
	; PRIOR APPLICATION NUMBER: US 60/404,249	
	; PRIOR FILING DATE: 2002-08-16	
	; PRIOR APPLICATION NUMBER: US 60/407,527	
	; PRIOR FILING DATE: 2002-08-28	
	; NUMBER OF SEQ ID NOS: 75	
	; SOFTWARE: Patentin version 3.2	
	; SEQ ID NO 30	
	; LENGTH: 2351	
	; TYPE: PRT	
	; ORGANISM: Homo sapiens	
	US-10-411-026-30	
	Query Match 99.9%; Score 12401; DB 12; Length 2351;	
	Best Local Similarity 99.9%; Pred. No. 0;	
	Matches 2349; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	1 MOELSTCFCLLRCPSATRRYYILGAVELSWDVWQSGLGELPVDARFPPRPVKSPFFN 60	
DB	1 MQEISTCFCLLRCPFSATRYYILGAVELSWDVWQSGLGELPVDARFPFRPKSPPFN 60	
QY	61 TSVVYKKTFLVFETDHLNFIAKRPRPMGLLPITQAEBYDTVTIWKXMASHPVSJHAV 120	
DB	61 TSVVYKKTFLVFETDHLNFIAKRPRPMGLLPITQAEBYDTVTIWKXMASHPVSJHAV 120	
QY	121 GVSYWKASGEAYDDOTSQRKKDKVPGGSHTVVWVLKENGPMSADPLCTLYSVLSH 180	
DB	121 GVSYWKASGEAYDDOTSQRKKDKVPGGSHTVVWVLKENGPMSADPLCTLYSVLSH 180	
QY	181 VDLVKDNLNSGLIGALIVCREGS LAKEQTQLHKFTILLFAVEDEGSKWSHSETKNLSMQDRD 240	
DB	181 VDLVKDNLNSGLIGALLVCREGSLAKEKXTQLHKFTILLFAVEDEGSKWSHSETKNLSMQDRD 240	
QY	241 AASARAWPQHWTWGYNVRNSPGLCIGHRKKSVMHWVGMTTPEVHSIFLEGHTFLVRNH 300	
DB	241 AASARAWPQHWTWGYNVRNSPGLCIGHRKKSVMHWVGMTTPEVHSIFLEGHTFLVRNH 300	
QY	301 RQASLEISPIFLTAAQTLLMDLGQFLSCHISSHQHDGMWAYKVYDVSCPEEPQRLRMKNNE 360	
DB	301 RQASLEISPIFLTAAQTLLMDLGQFLRSCHISSHQHDGMWAYKVYDVSCPEEPQRLMKNE 360	
QY	361 EADDYDDDITDSMDEVRFDDDNSPSFIQRSVAKKPKTWVHYTAABEEEDWDVAPLVLA 420	
DB	361 EADDYDDDITDSMDEVRFDDDNSPSFIQRSVAKKPKTWVHYTAABEEEDWDVAPLVLA 420	
QY	421 PDPRS VKSYDLANGPORICGRKYKRVFWAYTDDETFTREATCHSGII GPILI YGEVGDTL 480	
DB	421 PDPRS VKSQYLNGPORICGRKYKRVFWAYTDDETFTREACHESGI IGP ILI YGEVGD TL 480	
QY	481 LIIFNQASRPNYIPHGTTIDVRPLYSRRLPKGVKLKDFFILPGEIFPKYKWTVTVEDGP 540	

1561 GKVPFLRVATESSAKTPSKLLDPLAWDNHYGQIPEKEBWSQEKSPKTAFFKKDITLSL 1620
QY 1621 NACENHAIALNEGONKEEIVTWAKOQTERLCSQNPVLRKHORBITRTLOSQOE 1680
DB 1621 NACENHAIALNEGONKEEIVTWAKOQTERLCSQNPVLRKHORBITRTLOSQOE 1680
QY 1681 IDYDITISVEMKEKEDFDIYDEBENQSPRSFQKTRHYTAAVERLWDYGMSSSPVLRNR 1740
DB 1681 IDYDITISVEMKEKEDFDIYDEBENQSPRSFQKTRHYTAAVERLWDYGMSSSPVLRNR 1740
QY 1741 AOSGSVPQKVVFOEFTDGSSTQBLYRGELNEHGLGPGYIRAEVDENIMWTFNQASR 1800
DB 1741 AOSGSVPQKVVFOEFTDGSSTQBLYRGELNEHGLGPGYIRAEVDENIMWTFNQASR 1800
QY 1801 PYSFYSSLLSYBEDORQGAEPKKNFVKNETKTYFWKVOHMAPTKDPEDCAKAWYFSDV 1860
DB 1801 PYSFYSSLLSYBEDORQGAEPKKNFVKNETKTYFWKVOHMAPTKDPEDCAKAWYFSDV 1860
QY 1861 DLEKDVHSLIGLPLVCHTNTLPAHGRQVTVQEFALPFTIFDETKSWYFTENMERNCR 1920
DB 1861 DLEKDVHSLIGLPLVCHTNTLPAHGRQVTVQEFALPFTIFDETKSWYFTENMERNCR 1920
QY 1921 PCNIOMEDPTFKENYRFRHAINCYINDTLPLGLVMAQDQRIWYLLSGNSNENIHSYHPSGH 1980
DB 1921 PCNIOMEDPTFKENYRFRHAINCYINDTLPLGLVMAQDQRIWYLLSGNSNENIHSYHPSGH 1980
QY 1981 VFTVKEKKEKYNALNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGNSTILFLVYSNKC 2040
DB 1981 VFTVKEKKEKYNALNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGNSTILFLVYSNKC 2040
QY 2041 QTPGLMAGSHIRDPOTASQYQGWAPKLARLHYSGSINAWTKPEFFSWIKVDLLAPMII 2100
DB 2041 QTPGLMAGSHIRDPOTASQYQGWAPKLARLHYSGSINAWTKPEFFSWIKVDLLAPMII 2100
QY 2101 HGKTCGAZOKESSLYISQFIIMVSLDGKKWQTYRGNSTGTLWVPGNVDSSGIYHNIFN 2160
DB 2101 HGKTCGAZOKESSLYISQFIIMVSLDGKKWQTYRGNSTGTLWVPGNVDSSGIYHNIFN 2160
QY 2161 PPIIARYIRLHPTHYSIRSTRMELMGCNLSNCSPLGMESKAISDAQITASSYFTNMFA 2220
DB 2161 PPIIARYIRLHPTHYSIRSTRMELMGCNLSNCSPLGMESKAISDAQITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLOGRNAPQVNNPKEWLOVDFOQTKMKVGTGTVQVKSLLTSMYVKEFL 2280
DB 2221 TWSPSKARLHLOGRNAPQVNNPKEWLOVDFOQTKMKVGTGTVQVKSLLTSMYVKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVQGNODSFTPVVNSLDPLITRYLRHPQSWWHQIALRM 2340
DB 2281 ISSSQDGHQWTLFFQNGKVKVQGNODSFTPVVNSLDPLITRYLRHPQSWWHQIALRM 2340
QY 2341 EYLGCBAQDLY 2351
DB 2341 EYLGCBAQDLY 2351

RESULT 6

US-10-360-101-229
; Sequence 229, Application US/10360101
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; PRIOR FILING DATE: 2003-02-07
; PRIOR FILING DATE: 2003-02-07
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 229
; LENGTH: 2332
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of factor VIII
US-10-360-101-229
Query Match 99.1%; Score 12306; DB 15; Length 2332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2331; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 20 ATRRYVLCAVELSDWYQSDIGELPVDARPPRPVKSPPTNSVYVKKTLVEFTHLNF 79
DB 1 ATRRYVLCAVELSDWYQSDIGELPVDARPPRPVKSPPTNSVYVKKTLVEFTHLNF 60
QY 80 IAKRPPPMWGLLGPITQAEVYDVTVITLKNMASHPVSLHAGVSVKASGAEYDDQTSQ 139
DB 61 IAKRPPPMWGLLGPITQAEVYDVTVITLKNMASHPVSLHAGVSVKASGAEYDDQTSQ 120
QY 140 REKDDKVPFGSSTYVWVLKENGPMASDPLCTYSYLSHVLDLVKDLNSGLIGALLVCR 199
DB 121 REKDDKVPFGSSTYVWVLKENGPMASDPLCTYSYLSHVLDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKTOTLHKFILLFAVDEGKSWHSETKNSLMODRDAASARAKPKHETVNGYNR 259
DB 181 EGSIAKEKTOTLHKFILLFAVDEGKSWHSETKNSLMODRDAASARAKPKHETVNGYNR 240
QY 260 SLPLGICGRXSVVWHVIGMGTTPFVHSIFLEGHTFLVNRHQASLEISPIITLTAQTL 319
DB 241 SLPLGICGRXSVVWHVIGMGTTPFVHSIFLEGHTFLVNRHQASLEISPIITLTAQTL 300
QY 320 MDLGOFLLSCHISQHDGMEAYVKVDCPBEPPQIRKKNHEAEYDDDLTDSMDVVRP 379
DB 301 MDLGOFLLSCHISQHDGMEAYVKVDCPBEPPQIRKKNHEAEYDDDLTDSMDVVRP 360
QY 380 DDNSPSFIQIRSAKHPKTVWHYIAAEEDWDYAPLVLAPEDDRSYKSOYLNGPQIRG 439
DB 361 DDNSPSFIQIRSAKHPKTVWHYIAAEEDWDYAPLVLAPEDDRSYKSOYLNGPQIRG 420
QY 440 RYKXVRFMAYTDETFXTREAIQIHESGILGELLXGEVGTLLIIFKNQASRPYNIYHGI 499
DB 421 RYKXVRFMAYTDETFXTREAIQIHESGILGELLXGEVGTLLIIFKNQASRPYNIYHGI 480
QY 500 TDVRELYSRRLPKGVKHLKDPFLLPGEIFKXKWTVTVEDGPTKSDPRCLTRYSSFYVME 559
DB 481 TDVRELYSRRLPKGVKHLKDPFLLPGEIFKXKWTVTVEDGPTKSDPRCLTRYSSFYVME 540
QY 560 RDLASGLIGPLLI CYKESVDORGNQIMSKDNVILFSVDENRSWYLTENQRFLPNFAG 619
DB 541 RDLASGLIGPLLI CYKESVDORGNQIMSKDNVILFSVDENRSWYLTENQRFLPNFAG 600
QY 620 VOLEDPPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTPKH 679
DB 601 VOLEDPPEFQASNMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTPKH 660
QY 680 XWVYEDTLTLPPFGSETVMSNENPGLMILGCHNSDPENRGMTALLKYSSCDKNYGDYFE 739
DB 661 XWVYEDTLTLPPFGSETVMSNENPGLMILGCHNSDPENRGMTALLKYSSCDKNYGDYFE 720
QY 740 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTQKQFNATTPENDIEKTDPMFAHRTMP 799
DB 721 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTQKQFNATTPENDIEKTDPMFAHRTMP 780
QY 800 KIQNVSSDDLMLLRQSTPTPGLSLSDIQEAKYETFSDDPSPGAIDSNNSISEMTHFRPQ 859
DB 781 KIQNVSSDDLMLLRQSTPTPGLSLSDIQEAKYETFSDDPSPGAIDSNNSISEMTHFRPQ 840
QY 860 LHSQDMVFTPESGLOLRNEKLGTAATBLKLDPKVSSSTNNLITIPSDNLAAGTDN 919
DB 841 LHSQDMVFTPESGLOLRNEKLGTAATBLKLDPKVSSSTNNLITIPSDNLAAGTDN 900
QY 920 TSSLGPPGMVPHYDQDLOTTFLGKKSPLTBSGGPLSSENNDSKLLSGLMNSQESSW 979
DB 901 TSSLGPPGMVPHYDQDLOTTFLGKKSPLTBSGGPLSSENNDSKLLSGLMNSQESSW 960

QY	980	GKNSSTESGRIFKGRAGHGPALLTKDNLAFVYSISLLKTKNTKNSNATNRKTHIDGPSL	1039
Db	961	GKWSSTESGRIFKGRAGHGPALLTKONALFKVYSISLLKTKNTKNSNATNRKTHIDGPSL	1020
QY	1040	LEENSPSVQONILESDTEPKKVTPLIHRDMLMDKNATALLRNMGNKTTSSKNMVMQOK	1099
Db	1021	LEENSPSVQONILESDTEPKKVTPLIHRDMLMDKNATALLRNMGNKTTSSKNMVMQOK	1080
QY	1100	KEGPPIPPAQNDMDFKXMLFLPESARMIORTHGKNSLNSGGSPKQOLVSGPKSVRG	1159
Db	1081	KEGPPIPPAQNDMDFKXMLFLPESARMIORTHGKNSLNSGGSPKQOLVSGPKSVRG	1140
QY	1160	QNFLSEKNVVGKGEPTKDVGLKEMVPPSSRNLFILNLDNLHNNTHNQEKIIOEEIEK	1219
Db	1141	QNFLSEKNVVGKGEPTKDVGLKEMVPPSSRNLFILNLDNLHNNTHNQEKIIOEEIEK	1200
QY	1220	KETLLQENNVLPQIHVTGTXNFMKNLFLSTRQNVGSGYDCAVAPVLQDFRSLNDSNR	1279
Db	1201	KETLLQENNVLPQIHVTGTXNFMKNLFLSTRQNVGSGYDCAVAPVLQDFRSLNDSNR	1260
QY	1280	TKGHTAHFSKKEEENLEGLGNQTKOIVEKACTTRISNPTSQONFVFORSGKALKQREL	1339
Db	1261	TKGHTAHFSKKEEENLEGLGNQTKOIVEKACTTRISNPTSQONFVFORSGKALKQREL	1320
QY	1340	PLEETELEKRIIVDDTSTQWSKNMKHLPSTLTQIDYNEKKGKAITQSPDCLTRSHSI	1399
Db	1321	PLEETELEKRIIVDDTSTQWSKNMKHLPSTLTQIDYNEKKGKAITQSPDCLTRSHSI	1380
QY	1400	POANRSPUPIAKVSFPIRPIYITRVLFDQNSHLPAASTRKDSGVQOESSHFLQGAKK	1459
Db	1381	POANRSPUPIAKVSFPIRPIYITRVLFDQNSHLPAASTRKDSGVQOESSHFLQGAKK	1440
QY	1460	NNLSAILTLTLEMIGDQRBVGLSGTSATNSVYKVENLTKPDLPKTSKGVELLPKVHI	1519
Db	1441	NNLSAILTLTLEMIGDQRBVGLSGTSATNSVYKVENLTKPDLPKTSKGVELLPKVHI	1500
QY	1520	YOKDLFPETNSGPHLDLVEGSLLOQTEGALKWNEANRPCKVFLRVATESSAKTFSK	1579
Db	1501	YOKDLFPETNSGPHLDLVEGSLLOQTEGALKWNEANRPCKVFLRVATESSAKTFSK	1560
QY	1580	LLDPLANDNHVGTQIPKEEWSQOESPEKTAFAKDDIILSNACSNHAIAMINEGQNK	1639
Db	1561	LLDPLANDNHVGTQIPKEEWSQOESPEKTAFAKDDIILSNACSNHAIAMINEGQNK	1620
QY	1640	EIVTWAKQGRTERLCSQNPPLVKHOREIIRTTLSQDQEBIDYDDITISVEMKXEDFIY	1699
Db	1621	EIVTWAKQGRTERLCSQNPPLVKHOREIIRTTLSQDQEBIDYDDITISVEMKXEDFIY	1680
QY	1700	DEDENQSPRQKTRHYFTIAAVERLDYDGMSSSPHVLNRQAQSGVPPQPKVVFQFTD	1759
Db	1681	DEDENQSPRQKTRHYFTIAAVERLDYDGMSSSPHVLNRQAQSGVPPQPKVVFQFTD	1740
QY	1760	GSFTQPLRYGSLNEHLGLGPYIRAEVEDNIMVTFRQAQSPYFYSLSLISYEDQOQA	1819
Db	1741	GSFTQPLRYGSLNEHLGLGPYIRAEVEDNIMVTFRQAQSPYFYSLSLISYEDQOQA	1800
QY	1820	EPRKNFVKPNETKTYFWKQVHEMPTKDEDFCKAWAYFSDVLEKDVHSGLIGPLLCHT	1879
Db	1801	EPRKNFVKPNETKTYFWKQVHEMPTKDEDFCKAWAYFSDVLEKDVHSGLIGPLLCHT	1860
QY	1880	NTLNPAHGRQVTOEFALFTTIPDETQSWFTENMERNCRAPCNQIOMEDTFFKENYRFA	1939
Db	1861	NTLNPAHGRQVTOEFALFTTIPDETQSWFTENMERNCRAPCNQIOMEDTFFKENYRFA	1920
QY	1940	INGYIMDTPLGLVMAQOQIRWYLLSGNSNENIHSIHPSGHVFTVRKKEEYKXALNLYP	1999
Db	1921	INGYIMDTPLGLVMAQOQIRWYLLSGNSNENIHSIHPSGHVFTVRKKEEYKXALNLYP	1980
QY	2000	GVFETVEMLSKAGIRWVECLIGEBLHAGMSTLFLVYSNKCQTPGLMGASHIRDFQITAS	2059
Db	1981	GVFETVEMLSKAGIRWVECLIGEBLHAGMSTLFLVYSNKCQTPGLMGASHIRDFQITAS	2040
QY	2060	GOYGQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQAGAKTSSLYISQ	2119

Db	2041	GOYGQWAPKLARLHYSGINAWSTKEPFSWIKVDLLAPMIHGIKTQAGAKTSSLYISQ	2100
QY	2120	FIIMYSLDGKKWQYRGNGSTGTLMVFFGNVDSSGIGKHNIFNPPIIARVIRLHPTHSIRS	2179
Db	2101	FIIMYSLDGKKWQYRGNGSTGTLMVFFGNVDSSGIGKHNIFNPPIIARVIRLHPTHSIRS	2160
QY	2180	TLRBMELMGCDLNSCSMPLGMSKASIDAOITASSYFTNMPATWSPSKARLHLOQRSNAWR	2239
Db	2161	TLRBMELMGCDLNSCSMPLGMSKASIDAOITASSYFTNMPATWSPSKARLHLOQRSNAWR	2220
QY	2240	PQVNRPKEWLQVDFOKTMKVGTQGVKSLTSMYKKEPLISSODGHOWTLFQNGKV	2299
Db	2221	PQVNRPKEWLQVDFOKTMKVGTQGVKSLTSMYKKEPLISSODGHOWTLFQNGKV	2280
QY	2300	KVFCQNDSPFPVNSLDPPLLLTRYLRHQPQSVHQAIRMEVLGCEAQDLY	2351
Db	2281	KVFCQNDSPFPVNSLDPPLLLTRYLRHQPQSVHQAIRMEVLGCEAQDLY	2332
RESULT 7			
US-09-957-641-2			
; Sequence 2, Application US/09957641			
; Publication No. US20020182670A1			
; GENERAL INFORMATION:			
; APPLICANT: Emory University			
; TITLE OF INVENTION: MODIFIED FACTOR VIII			
; FILE REFERENCE: 75-00			
; CURRENT APPLICATION NUMBER: US/09/957,641			
; CURRENT FILING DATE: 2001-09-16			
; PRIOR APPLICATION NUMBER: US 60/234047			
; PRIOR FILING DATE: 2000-09-19			
; PRIOR APPLICATION NUMBER: US 60/236460			
; PRIOR FILING DATE: 2000-09-29			
; NUMBER OF SEQ ID NOS: 18			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 2			
; LENGTH: 2332			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-957-641-2			
Query Match 99.08; Score 12293; DB 9; Length 2332;			
Best Local Similarity 99.98; Pred. No. 0;			
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
QY	20	ATRRYILGAVELSDYMQSDLGELPVDARPPRPVPSFPFNTSVVYKTLFVETDHLFN	79
Db	1	ATRRYILGAVELSDYMQSDLGELPVDARPPRPVPSFPFNTSVVYKTLFVETDHLFN	60
QY	80	IAKPRPPMGLGPTIOAEVYDVTITLKNVSHPSLHAGVSYWKAEGAYDDQTSQ	139
Db	61	IAKPRPPMGLGPTIOAEVYDVTITLKNVSHPSLHAGVSYWKAEGAYDDQTSQ	120
QY	140	REKEDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	199
Db	121	REKEDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR	180
QY	200	EGSLAKEKTOHLKPIILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR	259
Db	181	EGSLAKEKTOHLKPIILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR	240
QY	260	SLPGLIGCHKSHVYHVGMTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL	319
Db	241	SLPGLIGCHKSHVYHVGMTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL	300
QY	320	MDLGQFLLSHSSHQDGMVAVYKVDSCPEBQPMKNNEAEYDDDLTDSMDVVR	379
Db	301	MDLGQFLLSHSSHQDGMVAVYKVDSCPEBQPMKNNEAEYDDDLTDSMDVVR	360
QY	380	DDNSPSFTQIRSAVAKKPKTWYVYTAABEEDWDYAPLVLPDDRYSKYQLNNGPQIRG	439
Db	361	DDNSPSFTQIRSAVAKKPKTWYVYTAABEEDWDYAPLVLPDDRYSKYQLNNGPQIRG	420

QY 440 RYKVVREMYATDETEKREAIQHESGILGPILYGEVGDITLIIIFKNOASRPYNIYPHGI 499
DB 421 RYKVVREMYATDETEKREAIQHESGILGPILYGEVGDITLIIIFKNOASRPYNIYPHGI 480
QY 500 TDVRLYSRRLLPKGVKHLKDPILPGEIFKYKWTVTVDGPTKSDPRCLTRYSSFVNE 559
DB 481 TDVRLYSRRLLPKGVKHLKDPILPGEIFKYKWTVTVDGPTKSDPRCLTRYSSFVNE 540
QY 560 RDLASGLIGPILICYKESVDQGNQOIMSDKRNVLFSVPDENRXYLTENQRFLENAG 619
DB 541 RDLASGLIGPILICYKESVDQGNQOIMSDKRNVLFSVPDENRXYLTENQRFLENAG 600
QY 620 VOLEDEPEFOASIMHSINGYVFDLSQLSVCLHEVAYWYILSCAQTDFLSVFFSGYTPKH 679
DB 601 VOLEDEPEFOASIMHSINGYVFDLSQLSVCLHEVAYWYILSCAQTDFLSVFFSGYTPKH 660
QY 680 KMVYEDTTLTFPFGSETVFMSENPGILWILGCHNSDFNRGMNTALLKVSSCDKNTGDYIE 739
DB 661 KMVYEDTTLTFPFGSETVFMSENPGILWILGCHNSDFNRGMNTALLKVSSCDKNTGDYIE 720
QY 740 DSYEDI SAVLLSKNAI PRSPSONSRHPSTKQFNATTIPENDIEKTDWPFAHRTMP 799
DB 721 DSYEDI SAVLLSKNAI PRSPSONSRHPSTKQFNATTIPENDIEKTDWPFAHRTMP 780
QY 800 KIQNVSSDILLMLLRQSPHPHGLSLDLOEAKYETFSDDPSGAIDSNLSLSEMTFRPQ 859
DB 781 KIQNVSSDILLMLLRQSPHPHGLSLDLOEAKYETFSDDPSGAIDSNLSLSEMTFRPQ 840
QY 860 LHSQDMVFTPSGLOLRNEXKLTGTAATELKKLDFKVSSTNNLITPSNLAAGTDN 919
DB 841 LHSQDMVFTPSGLOLRNEXKLTGTAATELKKLDFKVSSTNNLITPSNLAAGTDN 900
QY 920 TSSLGPPSMGVHDOLDITLFGKSSPPTESGGPLSLSEENNDKLLSGLMNSQESSW 979
DB 901 TSSLGPPSMGVHDOLDITLFGKSSPPTESGGPLSLSEENNDKLLSGLMNSQESSW 960
QY 980 GKNVSTESGRIFKGRAGHPALLTKONALFKVSI SLKTKNTSNNSAIPNRKTHIDGSL 1039
DB 961 GKNVSTESGRIFKGRAGHPALLTKONALFKVSI SLKTKNTSNNSAIPNRKTHIDGSL 1020
QY 1040 LIENSPSVWONILESTESKKTPTLJHDMMLDKATLRLHMSKNTTSSKNMVMQOK 1099
DB 1021 LIENSPSVWONILESTESKKTPTLJHDMMLDKATLRLHMSKNTTSSKNMVMQOK 1080
QY 1100 KEGPIPPDAQNDMFPFKMLFLPESARWIQRTGKNSLNSGOGPPKQVLSLGPESVBE 1159
DB 1081 KEGPIPPDAQNDMFPFKMLFLPESARWIQRTGKNSLNSGOGPPKQVLSLGPESVBE 1140
QY 1160 QNPLSEKKNVVGKGEFTKDVGLKENVFPSSRNLFITNLDNLHENNTHNQEKKI QEETEK 1219
DB 1141 QNPLSEKKNVVGKGEFTKDVGLKENVFPSSRNLFITNLDNLHENNTHNQEKKI QEETEK 1200
QY 1220 KETLLOENVVLQIHTVTCTKFMKNLFLSTRQNVGSGDYAYAPVLODFPSLNDSTNR 1279
DB 1201 KETLLOENVVLQIHTVTCTKFMKNLFLSTRQNVGSGDYAYAPVLODFPSLNDSTNR 1260
QY 1280 TKKHTAHFSGKGEENLEGLGNTQKQI VEKYACTTIRISNTSQONFVQTSRKALKQFRL 1339
DB 1261 TKKHTAHFSGKGEENLEGLGNTQKQI VEKYACTTIRISNTSQONFVQTSRKALKQFRL 1320
QY 1340 PLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
DB 1321 PLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
QY 1400 PQANRSPLEIAKVSSPPSRPIYLTVLQDNSSHLPAASYKKDQSGVQESSHFLOGAKK 1459
DB 1381 PQANRSPLEIAKVSSPPSRPIYLTVLQDNSSHLPAASYKKDQSGVQESSHFLOGAKK 1440
QY 1460 NNLSLAILTLEMTGDOREVGSLGTSATNSVYKVENTVLPKPDLPKTSKVELLPKVHI 1519
DB 1441 NNLSLAILTLEMTGDOREVGSLGTSATNSVYKVENTVLPKPDLPKTSKVELLPKVHI 1500

QY 1520 YQKDLFTPTSTNGSPGHLDLVEGSLLOCTEGAIKXNEANRPCKVPFLRVATESAKTPSK 1579
DB 1501 YQKDLFTPTSTNGSPGHLDLVEGSLLOCTEGAIKXNEANRPCKVPFLRVATESAKTPSK 1560
QY 1580 LLDPLANDNHVGTQIPKEEWKQESKPEKTAFAKCKDTILSNACESHAIAMINEGQNK 1639
DB 1561 LLDPLANDNHVGTQIPKEEWKQESKPEKTAFAKCKDTILSNACESHAIAMINEGQNK 1620
QY 1640 BIEVTWAKQGRTERLCSQNPVLRKHOREIRTRTLOSQOBEIDYDDTILSVEMKEDDPDIY 1699
DB 1621 BIEVTWAKQGRTERLCSQNPVLRKHOREIRTRTLOSQOBEIDYDDTILSVEMKEDDPDIY 1680
QY 1700 DEENQSPRSQKTRHYFIAAVERLNDYGMSSPHVLNRNQAQSGSVFPQKVVQFQETD 1759
DB 1681 DEENQSPRSQKTRHYFIAAVERLNDYGMSSPHVLNRNQAQSGSVFPQKVVQFQETD 1740
QY 1760 CSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVFRNQAASRPYFYSLSIYEEDQROGA 1819
DB 1741 CSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVFRNQAASRPYFYSLSIYEEDQROGA 1800
QY 1820 EPRKXEVKPNETKTYFWKVQHHVAPTDEFDCKAWAYESDVKDCHVSHGLIGPLLVCHT 1879
DB 1801 EPRKXEVKPNETKTYFWKVQHHVAPTDEFDCKAWAYESDVKDCHVSHGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVGEFALFFTFIDETKSWYPTENMERNCRAPONIEMEDPTFKENYRFA 1939
DB 1861 NTLNPAHGRQVTVGEFALFFTFIDETKSWYPTENMERNCRAPONIEMEDPTFKENYRFA 1920
QY 1940 INGYIMDTPLGLVNAQDORIEWYLLSMGSENHISIHESGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGYIMDTPLGLVNAQDORIEWYLLSMGSENHISIHESGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFEVTEMLPKAGIWRVECLIGEHLHAGMGTFLVYSNKQCTPLGWSAGHIRDPQITAS 2059
DB 1981 GVFEVTEMLPKAGIWRVECLIGEHLHAGMGTFLVYSNKQCTPLGWSAGHIRDPQITAS 2040
QY 2060 QYQOWAPKLARLHYSGSINAWSTKEPFSWIKVOLLAPMI IHGKTQGAROKFSSLYISQ 2119
DB 2041 QYQOWAPKLARLHYSGSINAWSTKEPFSWIKVOLLAPMI IHGKTQGAROKFSSLYISQ 2100
QY 2120 FIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGKKNINPPIIARVIRLHPHYSIRS 2179
DB 2101 FIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGKKNINPPIIARVIRLHPHYSIRS 2160
QY 2180 TIRMEIMGCDLNSCSPYLGMSKAISSDAQITASSYFTNMFWATWSPSKARLHLQGRSNAR 2239
DB 2161 TIRMEIMGCDLNSCSPYLGMSKAISSDAQITASSYFTNMFWATWSPSKARLHLQGRSNAR 2220
QY 2240 POVNNPKEWLQVDFQKTMKVTVGTTQGVKSLITSYVKEFLISSQDGHOWTLFPQNGKV 2299
DB 2221 POVNNPKEWLQVDFQKTMKVTVGTTQGVKSLITSYVKEFLISSQDGHOWTLFPQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDPPLLTLYRTHPOSVWHQIALRMEVLGCEAODLY 2351
DB 2281 KVFQGNQDSFTPVVNSLDPPLLTLYRTHPOSVWHQIALRMEVLGCEAODLY 2332

RESULT 8

US-10-187-319-2
; Sequence 2, Application US/10187319
; Publication No. US20030068785A1
; GENERAL INFORMATION:
; APPLICANT: Lolilar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESS: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 09/523,656
FILING DATE: 27-Aug-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26

ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: <unknown>
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:

ORGANISM: Homo sapiens
ISSUE TYPE: Liver
SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-10-187-319-2

Query Match 99.0%; Score 12293; DB 14; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

20 ATRRYVIGAVLSWDMYSDIGELPVDARFPRVPKSPBNTSVYKTLFVEFTDHLFN 79
1 ATRRYVIGAVLSWDMYSDIGELPVDARFPRVPKSPBNTSVYKTLFVEFTVHLFN 60

80 IAKRPPPMGLIGPTIQAEVYDVTVITLKNMASHPVSLHAGVSVYKASGAEYDDQTSQ 139
61 IAKRPPPMGLIGPTIQAEVYDVTVITLKNMASHPVSLHAGVSVYKASGAEYDDQTSQ 120

140 REKEDDKVPCGSHYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
121 REKEDDKVPCGSHYVWVLKENGPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180

200 EGSIAKEXTQTLHKFILLFAVFDGKSWHSETKNSLMODRDAASARAWPKMHTVNGYVNR 259
181 EGSIAKEXTQTLHKFILLFAVFDGKSWHSETKNSLMODRDAASARAWPKMHTVNGYVNR 240

260 SLPLGIGCHRSVYVHWITGMGTTPEVHISIFLEGHTFLVNRHQASLEISPTFLTAQTL 319
241 SLPLGIGCHRSVYVHWITGMGTTPEVHISIFLEGHTFLVNRHQASLEISPTFLTAQTL 300

320 MDLQGLFLSCHISSHODGMEAYKVDCSPEPQLRMKQNEAEADYDDDLTDSMDVVRP 379
301 MDLQGLFLSCHISSHODGMEAYKVDCSPEPQLRMKQNEAEADYDDDLTDSMDVVRP 360

380 DDNDSPSPQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLAPODRSVKSYKYNNGPQIRG 439
361 DDNDSPSPQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLAPODRSVKSYKYNNGPQIRG 420

440 RYKVKVRMAVTDFTFKTREAIQHESGLGPLLYGEVGDITLIIIFKNQASRPYNIYPHGI 499

421 RYKVKVRMAVTDFTFKTREAIQHESGLGPLLYGEVGDITLIIIFKNQASRPYNIYPHGI 480
500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 559
481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVAME 540
560 RDLASGLIGPLII CYKESVDORGNOIMDKKNVILFSVFDENRSWYLTENTQRFLEPAG 619
541 RDLASGLIGPLII CYKESVDORGNOIMDKKNVILFSVFDENRSWYLTENTQRFLEPAG 600
620 VQLEDPEFQASINHSINGYVFDLSQLSVCLHEVAYWVILSIGAQTDLSVFFSSYTKKH 679
601 VQLEDPEFQASINHSINGYVFDLSQLSVCLHEVAYWVILSIGAQTDLSVFFSSYTKKH 660
680 KMWVEDTLTLFPFSGETVFMSENPGLWILQCHNSDFENRGMTALLKVSSCDKNTGDIYE 739
661 KMWVEDTLTLFPFSGETVFMSENPGLWILQCHNSDFENRGMTALLKVSSCDKNTGDIYE 720
740 DSYEDISAYILSKNNAIEPRSPSQNSRHPSTROQFNATTIPENDIEKTDWFAHRTMP 799
721 DSYEDISAYILSKNNAIEPRSPSQNSRHPSTROQFNATTIPENDIEKTDWFAHRTMP 780
800 KIQNVSSDILLMLLRQSPTPHGLSLSDLOEAKYETFSDDPSGALDSNNSLSEMTHRPQ 859
781 KIQNVSSDILLMLLRQSPTPHGLSLSDLOEAKYETFSDDPSGALDSNNSLSEMTHRPQ 840
860 LHHSGDMVFTPESGLOLRINEKLGTTAATELKLDLFKVSSTSNLSTIPSDNLAAGTDN 919
841 LHHSGDMVFTPESGLOLRINEKLGTTAATELKLDLFKVSSTSNLSTIPSDNLAAGTDN 900
920 TSSLGPPSPVHYDSOLDTTLFGKSSDLTSGGPLSLSENNDSKLLSEGLMNSQESSW 979
901 TSSLGPPSPVHYDSOLDTTLFGKSSDLTSGGPLSLSENNDSKLLSEGLMNSQESSW 960
980 GKNVSTESGRLFKGKRAHGPAALLTKONALFKVSIILKTKNTKNNATNKHIDGPSL 1039
961 GKNVSTESGRLFKGKRAHGPAALLTKONALFKVSIILKTKNTKNNATNKHIDGPSL 1020
1040 LIENSPSVQWQNILBSTDTEFKKVTPLIHDRMLMDKNATRLNHNMSKNTTSSKNMEMVQOK 1099
1021 LIENSPSVQWQNILBSTDTEFKKVTPLIHDRMLMDKNATRLNHNMSKNTTSSKNMEMVQOK 1080
1100 KEGIP2PDAQNPDMSPFKMLFLPESARWIQTHGKNSINSCQGPSKQVSLGPEKSVEG 1159
1081 KEGIP2PDAQNPDMSPFKMLFLPESARWIQTHGKNSINSCQGPSKQVSLGPEKSVEG 1140
1160 QNFLSEKNKVVVGKGEFTKDVGLKEMVFPSSRNFLTNLMDLHNNTNHNQEKIQAESIEK 1219
1141 QNFLSEKNKVVVGKGEFTKDVGLKEMVFPSSRNFLTNLMDLHNNTNHNQEKIQAESIEK 1200
1220 KETLIQENVVLPOHTVTGTQKFMKNFLIISTRONVSGSYDGAAPVLQDPRSLNDSNR 1279
1201 KETLIQENVVLPOHTVTGTQKFMKNFLIISTRONVSGSYDGAAPVLQDPRSLNDSNR 1260
1280 TKKHTAFPSKGEENLEGLGNQTKQVVEKACATRISPNTSQONFVTOBSKRAKQFRL 1339
1261 TKKHTAFPSKGEENLEGLGNQTKQVVEKACATRISPNTSQONFVTOBSKRAKQFRL 1320
1340 PLEETELEKRIIIVDDTSTQWSKNMKHILTPSLTQIDYNEKEKGAIQSPISDCLTRSHSI 1399
1321 PLEETELEKRIIIVDDTSTQWSKNMKHILTPSLTQIDYNEKEKGAIQSPISDCLTRSHSI 1380
1400 POANRSLPIAKVSSPFIPIYILTRVLFDNNSHLLPAASVYRKDQSGVQSSSHFLQAKK 1459
1381 POANRSLPIAKVSSPFIPIYILTRVLFDNNSHLLPAASVYRKDQSGVQSSSHFLQAKK 1440
1460 NNLSLAIIITLMTGJOQREVGSLGTSATNSVYKVKVENTVLPKDLPKTSKVELLPKVI 1519
1441 NNLSLAIIITLMTGJOQREVGSLGTSATNSVYKVKVENTVLPKDLPKTSKVELLPKVI 1500
1520 YQKDLFTETSNGPSGHLDLVEGSLLOCTBGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579

Db 1501 YQKDLFPPTETSGSPGHLDLVEGSLLOCTEGAIKWEANRRKGVFLRVATSSAKTPSK 1560
Qy 1580 LLDPLAWNHYGTQIPKDEWSQKSPKTAARKKDDTILSNACSNHAIINNEGQNK 1639
Db 1561 LLDPLAWNHYGTQIPKDEWSQKSPKTAARKKDDTILSNACSNHAIINNEGQNK 1620
Qy 1640 BIEVTWAGQRTERLCSQNPVVKHQREIRITLQSQDEEIDYDDTISVENKKEEDFIY 1699
Db 1621 BIEVTWAGQRTERLCSQNPVVKHQREIRITLQSQDEEIDYDDTISVENKKEEDFIY 1680
Qy 1700 DEDENCSPRSFQKTRHYFIAAVERLWYGMSSSPHYVLRNRAQSGSVFQKVVQFQETD 1759
Db 1681 DEDENCSPRSFQKTRHYFIAAVERLWYGMSSSPHYVLRNRAQSGSVFQKVVQFQETD 1740
Qy 1760 GSFTQPLVRGELNEHLGLLGPVIRAEVDENIMVTFRNQSRPYSYSSLSIYVEDDQROGA 1819
Db 1741 GSFTQPLVRGELNEHLGLLGPVIRAEVDENIMVTFRNQSRPYSYSSLSIYVEDDQROGA 1800
Qy 1820 EPRKNFVKNETKTYFWKQVQHHNAPTDEFFCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
Db 1801 EPRKNFVKNETKTYFWKQVQHHNAPTDEFFCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
Qy 1880 NTLNPAHGRQVTVQEFALFFITFDTKWYFTENMERNCRAPCNQMDPTFKENYRFA 1939
Db 1861 NTLNPAHGRQVTVQEFALFFITFDTKWYFTENMERNCRAPCNQMDPTFKENYRFA 1920
Qy 1940 INGYIMDTLPGVMAQDQIRWYLLSMGSENHSHFSGHVFYVRKKEEYKXALYNYLP 1999
Db 1921 INGYIMDTLPGVMAQDQIRWYLLSMGSENHSHFSGHVFYVRKKEEYKXALYNYLP 1980
Qy 2000 GVFTVEMLPSKAGIWRVECLLGEHLHAGMSTFLVYNSKNCQTPGLMASGHIRDFQITAS 2059
Db 1981 GVFTVEMLPSKAGIWRVECLLGEHLHAGMSTFLVYNSKNCQTPGLMASGHIRDFQITAS 2040
Qy 2060 GOYGOWAPKARLHYSGSINASTKEPFSWIKVDLLAPMIHGKTOGARQKPFSSLYISQ 2119
Db 2041 GOYGOWAPKARLHYSGSINASTKEPFSWIKVDLLAPMIHGKTOGARQKPFSSLYISQ 2100
Qy 2120 FTIMYSLDGKKWQTRGNSGTGLVYFFGNVDSGKHNIFNPIIARIYRILHPTHYSIRS 2179
Db 2101 FTIMYSLDGKKWQTRGNSGTGLVYFFGNVDSGKHNIFNPIIARIYRILHPTHYSIRS 2160
Qy 2180 TURMELMGCDLNSCNPGLMESKALSDAQITASSYFTNNFATWSPSKALHLQGRSNWR 2239
Db 2161 TURMELMGCDLNSCNPGLMESKALSDAQITASSYFTNNFATWSPSKALHLQGRSNWR 2220
Qy 2240 POWNPKWLOYDFOKTMAVTGVTQGVKSLLTSMVYKBFLLISSODGHQWTLFQNGKV 2299
Db 2221 POWNPKWLOYDFOKTMAVTGVTQGVKSLLTSMVYKBFLLISSODGHQWTLFQNGKV 2280
Qy 2300 KVFQGNQDSFTPVNSLDPPLTRYLRIHFQSWVHQIALRMEVLCGEAQDLY 2351
Db 2281 KVFQGNQDSFTPVNSLDPPLTRYLRIHFQSWVHQIALRMEVLCGEAQDLY 2332

RESULT 9
US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: J.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26

; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-131-510A-2

Query Match 99.0%; Score 12293; DB 14; Length 2332;
Best local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 20 ATRRYYLGAVELSWDYNQSDIGELPVDARPPRPVKFPFNTSVYKTLFVEFTDHLFN 79
Db 1 ATRRYYLGAVELSWDYNQSDIGELPVDARPPRPVKFPFNTSVYKTLFVEFTVHLFN 60

Qy 80 IAKRPMPMGLLGTTIQAENVDTNITLKNASHPVSLHAGVSYWKASEGAYDDQTSQ 139
Db 61 IAKRPMPMGLLGTTIQAENVDTNITLKNASHPVSLHAGVSYWKASEGAYDDQTSQ 120

Qy 140 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 199
Db 121 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKOLNSGLIGALLVCR 180

Qy 200 EGSIAKAKTQTLHKPILLPAVFDGKSWHSTKSLMODRDAASARAWPKMHTVNGYNR 259
Db 181 EGSIAKAKTQTLHKPILLPAVFDGKSWHSTKSLMODRDAASARAWPKMHTVNGYNR 240

Qy 260 SLPLGLIGCHRSVYVWVGMGTTPRVHSIFLEGHTFLVNRHQASLEISPIITLTAQTL 319
Db 241 SLPLGLIGCHRSVYVWVGMGTTPRVHSIFLEGHTFLVNRHQASLEISPIITLTAQTL 300

Qy 320 MDLGOFLLSCHISSHQHDMGMEAYVKVDCSPEPQLRMKNNEAEEDYDDLDSEMDVYRF 379
Db 301 MDLGOFLLSCHISSHQHDMGMEAYVKVDCSPEPQLRMKNNEAEEDYDDLDSEMDVYRF 360

Qy 380 DDNSPSPTQIRSVAKHPKTVWYIAAEEEDWDYAPVLAPDDRSYSQSYLNNGPQIRG 439
Db 361 DDNSPSPTQIRSVAKHPKTVWYIAAEEEDWDYAPVLAPDDRSYSQSYLNNGPQIRG 420

Qy 440 RYKXKVRFMAYTDEFKTRERAIQHSGLGPLLYGEVGDITLLIFKNQASRPYNIYPHG 499
Db 421 RYKXKVRFMAYTDEFKTRERAIQHSGLGPLLYGEVGDITLLIFKNQASRPYNIYPHG 480

Qy 500 TDVRPLYSRRLPKGVKHLKDPFIPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 559
Db 481 TDVRPLYSRRLPKGVKHLKDPFIPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 540

Qy 560 RDLASGLIGPLLIICYKESVDQGNQIMSKDNVILFSVFDENRSKYLNTORFLPNPAG 619
Db 541 RDLASGLIGPLLIICYKESVDQGNQIMSKDNVILFSVFDENRSKYLNTORFLPNPAG 600

Qy 620 VOLEDPPEFQASNIHNSINGYVDFSQLSVCLHEVAYWILSGAQTDFLSVFFSGYTFKH 679
Db 601 VOLEDPPEFQASNIHNSINGYVDFSQLSVCLHEVAYWILSGAQTDFLSVFFSGYTFKH 660

Qy 680 KMVYEDTLTLFPFSGETVPMGMENPGLWILGCHNSDFNRNGMTALLKVSQCDKNTGYE 739
Db 661 KMVYEDTLTLFPFSGETVPMGMENPGLWILGCHNSDFNRNGMTALLKVSQCDKNTGYE 720

Qy 740 DSYEDISAYLLSKNNAIEPRFSQNSRHPSTOKOFNATITPENDIEKTDPAHRTMP 799
Db 721 DSYEDISAYLLSKNNAIEPRFSQNSRHPSTOKOFNATITPENDIEKTDPAHRTMP 780

Qy 800 KIQNVSSSSDLLMLRQSPHPHGLSLDQEAKEYTFSDPSPGADSNNSLSEMTFRPQ 859
Db 781 KIQNVSSSSDLLMLRQSPHPHGLSLDQEAKEYTFSDPSPGADSNNSLSEMTFRPQ 840

QY 860 LHHSGDMVFTPESSGLOLRINEKLGTTAATELKKLDKFKVSSSTNNLISTIPSDNLAAGTDN 919
DB 841 LHHSGDMVFTPESSGLOLRINEKLGTTAATELKKLDKFKVSSSTNNLISTIPSDNLAAGTDN 900
QY 920 TSSLGPPMPVHYDPSQDITTLFGKKSPLTBSGGLSISEENNDPSKLESGLMNSQSSW 979
DB 901 TSSLGPPMPVHYDPSQDITTLFGKKSPLTBSGGLSISEENNDPSKLESGLMNSQSSW 960
QY 980 GKNVSSTESGLFGKKAHGPALLTKONALFKVSIILKTKNTKNSATRKTHIDGPSL 1039
DB 961 GKNVSSTESGLFGKKAHGPALLTKONALFKVSIILKTKNTKNSATRKTHIDGPSL 1020
QY 1040 LIENSPSVWQNILEDSTEFKVTPLIHDRMLMDKNATRLNEMSNKTTSSKNEMVQOK 1099
DB 1021 LIENSPSVWQNILEDSTEFKVTPLIHDRMLMDKNATRLNEMSNKTTSSKNEMVQOK 1080
QY 1100 KEGPIPPDAQNPDMSFFROMLFLPSARWIOQTHGKNSLNSQGPSKQLVSLGPEKSV 1159
DB 1081 KEGPIPPDAQNPDMSFFROMLFLPSARWIOQTHGKNSLNSQGPSKQLVSLGPEKSV 1140
QY 1160 QNFUSEKNKVVVGKGETKOVGLKEMVFPSSRNFLTKNLDNLHENNTHNQEKKEQBI 1219
DB 1141 QNFUSEKNKVVVGKGETKOVGLKEMVFPSSRNFLTKNLDNLHENNTHNQEKKEQBI 1200
QY 1220 KETLIQENVVLPOHTVTGTGNFMKNFLIISTRQNVGSGYDGAAPVLQDPRSLNDS 1279
DB 1201 KETLIQENVVLPOHTVTGTGNFMKNFLIISTRQNVGSGYDGAAPVLQDPRSLNDS 1260
QY 1280 TKKHTAHFSKGBEENLEGLNQTKQIVKVIACATRTISPNTSQNFVQTSKRAKQFRL 1339
DB 1261 TKKHTAHFSKGBEENLEGLNQTKQIVKVIACATRTISPNTSQNFVQTSKRAKQFRL 1320
QY 1340 PLEETELEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKXKGAITQSPISDCITRSHSI 1399
DB 1321 PLEETELEKRIIVDDTSTQSKNKHITPSTLTQIDYNEKXKGAITQSPISDCITRSHSI 1380
QY 1400 POANRSLPIAKVSSPFSIRPIYLRVLFQDNSSHLPAASRYKKDQSGVQSSSHFLOGAKK 1459
DB 1381 POANRSLPIAKVSSPFSIRPIYLRVLFQDNSSHLPAASRYKKDQSGVQSSSHFLOGAKK 1440
QY 1460 NNLSLAITLTMTCQDQREVSLGTSAINSVYTKVENVTLPKDPLTKSGKVLLPKVHI 1519
DB 1441 NNLSLAITLTMTCQDQREVSLGTSAINSVYTKVENVTLPKDPLTKSGKVLLPKVHI 1500
QY 1520 YQKDLFFPTEISNGSPGHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESAKTPSK 1579
DB 1501 YQKDLFFPTEISNGSPGHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESAKTPSK 1560
QY 1580 LLDPLAWDNHYGTQIPKEBWKQSKSPEKTAPEKCKDTILSLNACESNHAIAAINEGQNK 1639
DB 1561 LLDPLAWDNHYGTQIPKEBWKQSKSPEKTAPEKCKDTILSLNACESNHAIAAINEGQNK 1620
QY 1640 EIEVTWAKQRTSLCSQNPVLKGHOREITRTTLOSQDEIDYDDITSIYEMKEDFDIY 1699
DB 1621 EIEVTWAKQRTSLCSQNPVLKGHOREITRTTLOSQDEIDYDDITSIYEMKEDFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFIAAVERLMDYGMSSPHVLRNRAQSGVPQPKVVFQBF 1759
DB 1681 DEDENQSPRSFQKTRHYFIAAVERLMDYGMSSPHVLRNRAQSGVPQPKVVFQBF 1740
QY 1760 GSTQPLRYGELNEHGLGLOPYTRAENVEDIMVTRNQAQSRPYSFYSSLLSYBEDQOQA 1819
DB 1741 GSTQPLRYGELNEHGLGLOPYTRAENVEDIMVTRNQAQSRPYSFYSSLLSYBEDQOQA 1800
QY 1820 EPRKNFYKPNETKTYFWKVOHHMAPTDEDFCKAWAYFSVDLEKOVHSGLIGELLVCHT 1879
DB 1801 EPRKNFYKPNETKTYFWKVOHHMAPTDEDFCKAWAYFSVDLEKOVHSGLIGELLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFTTIFDETSKWYFTENWERNCRAPCNQIOMEDPTFKENYRFA 1939
DB 1861 NTLNPAHGRQVTVQEFALFTTIFDETSKWYFTENWERNCRAPCNQIOMEDPTFKENYRFA 1920

QY 1940 INGVIYMDTLFGLVMAQDQRIRWYLLSMGSENTHSIHPSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGVIYMDTLFGLVMAQDQRIRWYLLSMGSENTHSIHPSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVPETVEMLPKAGIMRVECLIGEHLMAGMSTLFLVYSNKCQTPPLGVASGHIRDFOITAS 2059
DB 1981 GVPETVEMLPKAGIMRVECLIGEHLMAGMSTLFLVYSNKCQTPPLGVASGHIRDFOITAS 2040
QY 2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGAOKFSSLYISQ 2119
DB 2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGAOKFSSLYISQ 2100
QY 2120 FIWYSLDGKKWOTYRGNSGTLMVFPNGVDSGKIKENIENPPIIARYIRLHPHYIRS 2179
DB 2101 FIWYSLDGKKWOTYRGNSGTLMVFPNGVDSGKIKENIENPPIIARYIRLHPHYIRS 2160
QY 2180 TLRMELMGCDLNSCMLPGVESKAISDAQITASSYFTNMFATWSPSKARLHQLGRSNAR 2239
DB 2161 TLRMELMGCDLNSCMLPGVESKAISDAQITASSYFTNMFATWSPSKARLHQLGRSNAR 2220
QY 2240 POVNNPKEWLOVDPQKTMKVTGVTTOGVKSLLTSMYKFEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 POVNNPKEWLOVDPQKTMKVTGVTTOGVKSLLTSMYKFEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KVFQGNODSTPVVNSLDPPLLTRYLRHPQSVWHQIALRMEVLGCEAQDLY 2351
DB 2281 KVFQGNODSTPVVNSLDPPLLTRYLRHPQSVWHQIALRMEVLGCEAQDLY 2332

RESULT 10

US-10-445-235-2
; Sequence 2, Application: US/10445235
; Publication No. US2004000570A1
; GENERAL INFORMATION:
; APPLICANT: Katherine A. High
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: CHOP0176
; CURRENT APPLICATION NUMBER: US/10/445,235
; PRIOR FILING DATE: 2003-05-22
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-445-235-2

Query Match 99.0%; Score 12293; DB 15; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 20 ATRRYVLGAVELSWDMQSDLGELPVDARPPPKSPFPNTGVYVKTLFVEFTDHLFN 79
DB 1 ATRRYVLGAVELSWDMQSDLGELPVDARPPPKSPFPNTGVYVKTLFVEFTDHLFN 60
QY 80 IAKPRPPMGLLQPTTQAEVYDTVWITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPMGLLQPTTQAEVYDTVWITLKNMASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVPPGGSHYVWQVLKENGSMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVPPGGSHYVWQVLKENGSMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKQTTLHKFTILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKQHTWGVNR 259
DB 181 EGSIAKEKQTTLHKFTILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKQHTWGVNR 240
QY 260 SIPGLLGCGRKSVYHVGNGTTPPEVHSIFLEGHTFLVRNHQASLEISPTFLTAQTL 319

Db 241 SLPLGLIGHRKSVYVHVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLL 300
Qy 320 MDLGGFLLSCHISSHOHDMAYVKVDCSPBPQLAMKNREAEEDYDDDLTDSMDVYRF 379
Db 301 MDLGGFLLSCHISSHOHDMAYVKVDCSPBPQLAMKNREAEEDYDDDLTDSMDVYRF 360
Qy 380 DDNSPFSFIQRSVAKKPKTWVHIAAAEEEDWDYAPLAPADDRSYSQYLNNGPQIG 439
Db 361 DDNSPFSFIQRSVAKKPKTWVHIAAAEEEDWDYAPLAPADDRSYSQYLNNGPQIG 420
Qy 440 RYKXVRFMAYTDEFTKTRAEIOHESGILGPLLYGEVGTLLIIFKNOASRPYNIYPHGI 499
Db 421 RYKXVRFMAYTDEFTKTRAEIOHESGILGPLLYGEVGTLLIIFKNOASRPYNIYPHGI 480
Qy 500 TDVRLYSRRLPKGVKHLKDFPILGELFKYKWTYVEDGPTKSDPRCLTRYSSFYVME 559
Db 481 TDVRLYSRRLPKGVKHLKDFPILGELFKYKWTYVEDGPTKSDPRCLTRYSSFYVME 540
Qy 560 RDLASGLIGPLIYCVKESVDQRGNOIMSDKRNVLFSVPDENRSWYLTENIQRPENPAG 619
Db 541 RDLASGLIGPLIYCVKESVDQRGNOIMSDKRNVLFSVPDENRSWYLTENIQRPENPAG 600
Qy 620 VOLEPPEOASNMHSINGYVFDLSQLSVCLHEVAYWYILS'GAOTDFLSVFSGYTFKH 679
Db 601 VOLEPPEOASNMHSINGYVFDLSQLSVCLHEVAYWYILS'GAOTDFLSVFSGYTFKH 660
Qy 680 KMVYEDTLTLFPFSGETVMSMENPCLWLGLCHNSDFRNGMTALLKYSSCDKNTGDYIE 739
Db 661 KMVYEDTLTLFPFSGETVMSMENPCLWLGLCHNSDFRNGMTALLKYSSCDKNTGDYIE 720
Qy 740 DSYEDISAYLLSKNNAIBPSPSONSRHPESTRQKQFNATTIPENDIEKTDPMFAHRTMP 799
Db 721 DSYEDISAYLLSKNNAIBPSPSONSRHPESTRQKQFNATTIPENDIEKTDPMFAHRTMP 780
Qy 800 KTONVSSDILLMLLQSPTPHGLSLSDIOEAKYETFSDDPSGALDSNNLSSEMTHPRPQ 859
Db 781 KTONVSSDILLMLLQSPTPHGLSLSDIOEAKYETFSDDPSGALDSNNLSSEMTHPRPQ 840
Qy 860 LHSQDMVFTPESGLQRLNEXLGTAAATELKLDFKVSSTSNLIISTIPSNLAAGTDN 919
Db 841 LHSQDMVFTPESGLQRLNEXLGTAAATELKLDFKVSSTSNLIISTIPSNLAAGTDN 900
Qy 920 TSSLGPPSPMEVHYDQDLDTLFGKXSSPLTESGGPLSLEENNDKLLSGLMNSQESSW 979
Db 901 TSSLGPPSPMEVHYDQDLDTLFGKXSSPLTESGGPLSLEENNDKLLSGLMNSQESSW 960
Qy 980 GKNVSTESGRILPKGRAHPALLTKONALFKVISISLLKNTKTSNNSATNRKTHIDGFSL 1039
Db 961 GKNVSTESGRILPKGRAHPALLTKONALFKVISISLLKNTKTSNNSATNRKTHIDGFSL 1020
Qy 1040 LIENSPSVWQNIILESTPEKKVTPLIHDMMLDKNATALLRNMKNKTTSSKNMWMVOOK 1099
Db 1021 LIENSPSVWQNIILESTPEKKVTPLIHDMMLDKNATALLRNMKNKTTSSKNMWMVOOK 1080
Qy 1100 KEGPIPPDAQNDPMGFFKMLFPEGARWQTHRGXNSLNSGCGPPKQIVSLGPKSVSEG 1159
Db 1081 KEGPIPPDAQNDPMGFFKMLFPEGARWQTHRGXNSLNSGCGPPKQIVSLGPKSVSEG 1140
Qy 1160 QNPLSEKNKVVYKGEPTKDVGLKENVPPSSRNLFITNLDNHNHNTHQEKIOEIEBK 1219
Db 1141 QNPLSEKNKVVYKGEPTKDVGLKENVPPSSRNLFITNLDNHNHNTHQEKIOEIEBK 1200
Qy 1220 KETLQENNVLPQIHVTGTQKFMKNLFLISTRQNVESYDQAYAPVLPDFSLNDSINR 1279
Db 1201 KETLQENNVLPQIHVTGTQKFMKNLFLISTRQNVESYDQAYAPVLPDFSLNDSINR 1260
Qy 1280 TKKTAHFKKGEENLEGLNQTKOIVBKYACTTRISPNTSQQNFVTOQRKRALKQFRL 1339
Db 1261 TKKTAHFKKGEENLEGLNQTKOIVBKYACTTRISPNTSQQNFVTOQRKRALKQFRL 1320
Qy 1340 PLEETELESRIIVDDTSTOWSNKMHLPSTLTQIDYNEKEKGAITQSPSLCSTRSHI 1399
Db 1321 PLEETELESRIIVDDTSTOWSNKMHLPSTLTQIDYNEKEKGAITQSPSLCSTRSHI 1380

Qy 1400 PQANRSLPIAKVSSPBIPIYILTRVLPODNSSHLPAASVYRKXDSGVQBSHSLQGAKK 1459
Db 1381 PQANRSLPIAKVSSPBIPIYILTRVLPODNSSHLPAASVYRKXDSGVQBSHSLQGAKK 1440
Qy 1460 NNLSLAILTLEMTGDQREVSGSLGTSATNSVTYKKEVENTVLPKDLPKTSGKVELLPKVI 1519
Db 1441 NNLSLAILTLEMTGDQREVSGSLGTSATNSVTYKKEVENTVLPKDLPKTSGKVELLPKVI 1500
Qy 1520 YOKLFPFTETNSGPHGLDLVEGSLLOSTEGAI'KWNANRPGKVPFFLRVATESAKTPSK 1579
Db 1501 YOKLFPFTETNSGPHGLDLVEGSLLOSTEGAI'KWNANRPGKVPFFLRVATESAKTPSK 1560
Qy 1580 LLDPLANDNHVGTQIPKEEWKSQESKPEKTAFAKKDITLSNACESHAI'AINNEGQNK 1639
Db 1561 LLDPLANDNHVGTQIPKEEWKSQESKPEKTAFAKKDITLSNACESHAI'AINNEGQNK 1620
Qy 1640 EIEVTWAKQGRTERLCSQNPPLVKRHOEITRTTLOSDQOEIDYDDTISVEMKKEDEFIY 1699
Db 1621 EIEVTWAKQGRTERLCSQNPPLVKRHOEITRTTLOSDQOEIDYDDTISVEMKKEDEFIY 1680
Qy 1700 DEDNQSPRSQKTRHYFIAAVERLWDYGVSSPHVLNRBAQSGSVQPKKVPQBEFTD 1759
Db 1681 DEDNQSPRSQKTRHYFIAAVERLWDYGVSSPHVLNRBAQSGSVQPKKVPQBEFTD 1740
Qy 1760 GSFTQPLVYRGELNHLGLLGPYIAEVEDNIMVTFRNQAASRPYFYSSLSIYEBEDQOGA 1819
Db 1741 GSFTQPLVYRGELNHLGLLGPYIAEVEDNIMVTFRNQAASRPYFYSSLSIYEBEDQOGA 1800
Qy 1820 EPRXNFVKPNETKTYFKVQHHMAPTKDEPCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
Db 1801 EPRXNFVKPNETKTYFKVQHHMAPTKDEPCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
Qy 1880 NTLPAHGRQVTVGEFALPFTIFDETSKWYFTEENMERNCRAPONTOMEDPTKENYRFA 1939
Db 1861 NTLPAHGRQVTVGEFALPFTIFDETSKWYFTEENMERNCRAPONTOMEDPTKENYRFA 1920
Qy 1940 INGYIMDTPLGLVWAQDQIRWYLLSMGSENHIIHPSGHVFTVRKKEEYKMALYNLYP 1999
Db 1921 INGYIMDTPLGLVWAQDQIRWYLLSMGSENHIIHPSGHVFTVRKKEEYKMALYNLYP 1980
Qy 2000 GVFETVEMLPKAGIWRVECLIGEHLHAGMTLFLVYSNKKCQTPLGWASGHIRDFQITAS 2059
Db 1981 GVFETVEMLPKAGIWRVECLIGEHLHAGMTLFLVYSNKKCQTPLGWASGHIRDFQITAS 2040
Qy 2060 QOYQOWAPKRLARLHYSGSINAWSTKEPESWIKVDLAPMIIHGKTQGAROKFSSLYISQ 2119
Db 2041 QOYQOWAPKRLARLHYSGSINAWSTKEPESWIKVDLAPMIIHGKTQGAROKFSSLYISQ 2100
Qy 2120 FIIMYSLDGKXQVYRGNSTGTLMVFFGNVDSSGTHKNIIPNPPIIARIYIRLHPHYSIRS 2179
Db 2101 FIIMYSLDGKXQVYRGNSTGTLMVFFGNVDSSGTHKNIIPNPPIIARIYIRLHPHYSIRS 2160
Qy 2180 TLRLMELMGCDLNSCMLPLGMSKAI'ISDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2239
Db 2161 TLRLMELMGCDLNSCMLPLGMSKAI'ISDAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2220
Qy 2240 PQVNNPKEWLOVDFOKTKMKTGVTTQGVKSLLTSMVYKEPLISSODGHQWTLFQNGKV 2299
Db 2221 PQVNNPKEWLOVDFOKTKMKTGVTTQGVKSLLTSMVYKEPLISSODGHQWTLFQNGKV 2280
Qy 2300 KVFQGNQDSFPPVNSLDPPLLTLYR'IHPOSVWHQIALRMEVLGCEAQDLY 2351
Db 2281 KVFQGNQDSFPPVNSLDPPLLTLYR'IHPOSVWHQIALRMEVLGCEAQDLY 2332

RESULT 11
US-10-239-498A-2
; Sequence 2, Application US/10239498A
; Publication No. US2004002333A1
; GENERAL INFORMATION:
; APPLICANT: Hauser, Charlotte
; APPLICANT: Horster, Andrea

Query Match	99.0%;	Score 12293;	DB 16;	Length 2332;
Best Local Similarity	99.9%;	Pred No. 0;		
Matches 2329;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0
QY	20	ATRYRYAGAVELSDYMQSDLGELFVDARFPFRVPKSPFNSTVYVYKKTLEVFETDHLFN	79	
Db	1	ATRYRYLGAVELSDYMQSDLGELFVDARFPFRVPKSPFNSTVYVYKKTLEVFETDHLFN	60	
QY	80	IAPRPFWMLLGGTTIQAEVYDVTWITLKNMASHPVSLHVGVSYWKASGAEVDDQTSQ	139	
Db	61	IAPRPFWMLLGGTTIQAEVYDVTWITLKNMASHPVSLHVGVSYWKASGAEVDDQTSQ	120	
QY	140	REKEDDVFFGGSTYVQVLKENGPMASDPLCLTYSYLSHVDVLKDLNSGLIGALLVCR	199	
Db	121	REKEDDVFFGGSTYVQVLKENGPMASDPLCLTYSYLSHVDVLKDLNSGLIGALLVCR	180	
QY	200	EGSLAKETQTLHKPFILLFAVDFDEKSWHSETKNSLMQDRDAASARAWPKVHTVGVYNR	259	
Db	181	EGSLAKETQTLHKPFILLFAVDFDEKSWHSETKNSLMQDRDAASARAWPKVHTVGVYNR	240	
QY	260	SLPGLIGCHRKSVVHWYVIGMGTTEVHSIFLHGHTFLVNRHQASLEISPTIFLTAQTLL	319	
Db	241	SLPGLIGCHRKSVVHWYVIGMGTTEVHSIFLHGHTFLVNRHQASLEISPTIFLTAQTLL	300	
QY	320	MDLQCFLLSCHISSHQHDGMEAYVKVDSCEPPEPOLRYKKNSEAEYDDDLTDSEMDVVRP	379	
Db	301	MDLQCFLLFCHISSHQHDGMEAYVKVDSCEPPEPOLRYKKNSEAEYDDDLTDSEMDVVRP	360	
QY	380	DDDNPSPIQIRSVAKHPKTTWVHYIAEEDWDYAPLVLPADDRSYKSQYLNNGPQIG	439	
Db	361	DDDNPSPIQIRSVAKHPKTTWVHYIAEEDWDYAPLVLPADDRSYKSQYLNNGPQIG	420	
QY	440	RKYKRVFMAYTDTFTKTRAIQESGILGPLLYGEVGTLLIIFKNQASPYNIYPGKI	499	
Db	421	RKYKRVFMAYTDTFTKTRAIQESGILGPLLYGEVGTLLIIFKNQASPYNIYPGKI	480	
QY	500	TDVREPLYSRRLPKGVKHLKDPFILLPGEIFYKWTVTVEDGTPKSDPRCLTRYSSFVME	559	
Db	481	TDVREPLYSRRLPKGVKHLKDPFILLPGEIFYKWTVTVEDGTPKSDPRCLTRYSSFVME	540	
QY	560	RDLASGLIGPLLI CYKESVDORGQIMSDKRNVLFSVPDENRSWYLTENIQRFLPNPAG	619	
Db	541	RDLASGLIGPLLI CYKESVDORGQIMSDKRNVLFSVPDENRSWYLTENIQRFLPNPAG	600	
QY	620	VQLEDPRPQANIMHSINGVYFDSIQLSVCLHEVAYWILSIGAOTDFLSVFFSGYTPKH	679	
Db	601	VQLEDPRPQANIMHSINGVYFDSIQLSVCLHEVAYWILSIGAOTDFLSVFFSGYTPKH	660	
QY	680	KWYVEDTLTLFPFGSETVFMSENPGLWILGCHNSDFNRGMGTALLKVSSCDKNTGDIYE	739	
Db	661	KWYVEDTLTLFPFGSETVFMSENPGLWILGCHNSDFNRGMGTALLKVSSCDKNTGDIYE	720	
QY	740	DSYEDISAYLSKNAIEBPRGFSQNSRHPSTQKQFNATTTPENDIEKTDPMFAHRTDMP	799	
Db	721	DSYEDISAYLSKNAIEBPRGFSQNSRHPSTQKQFNATTTPENDIEKTDPMFAHRTDMP	780	

```

1861 NTLNPAFGQVTVQBEFALFETITDETQSWYFTENMERNCRAPCNIQWEDPTKENVRFHA 1920
1940 INGYIMDTLPGLVMAODQORIRWLLMGSGNENITHESGHVETVRKKEKBYKMAVNLYP 1999
1921 INGYIMDTLPGLVMAODQORIRWLLMGSGNENITHESGHVETVRKKEKBYKMAVNLYP 1980
2000 GVFETVEMLPKAGIWRVCLAGEHLHAGMSTLFLVYSNKCOTPLGMAWGHIRDFOITAS 2059
1981 GVFETVEMLPKAGIWRVCLAGEHLHAGMSTLFLVYSNKCOTPLGMAWGHIRDFOITAS 2040
2060 GOYGOWAPKLARLHYGSGINAMSTKPFPSWIKVDLLAPMIIHGIKTQGARQKPSYIISQ 2119
2041 GOYGOWAPKLARLHYGSGINAMSTKPFPSWIKVDLLAPMIIHGIKTQGARQKPSYIISQ 2100
2120 FIIMYSLDGKKQWYRGNSGTILMWFFGVNDSSGIKHNIENPPIIARYIRLHPTHYSIRS 2179
2101 FIIMYSLDGKKQWYRGNSGTILMWFFGVNDSSGIKHNIENPPIIARYIRLHPTHYSIRS 2160
2180 TLRMELMGDLNCSXPLMGESKAISDAQITASSYFTNMFMATWSPSKARLHLQGRSNAR 2239
2161 TLRMELMGDLNCSXPLMGESKAISDAQITASSYFTNMFMATWSPSKARLHLQGRSNAR 2220
2240 PQVNNPEKWLQVDFQKTMKVTGVTTCGVKSLLTSMYVKZFLISSQDGHQWTLFFQNGKV 2299
2221 PQVNNPEKWLQVDFQKTMKVTGVTTCGVKSLLTSMYVKZFLISSQDGHQWTLFFQNGKV 2280
2300 KVFQGNQDSFTPVNSLDPPLTRYLRIHPQSWHVOIALRMEVLGCEAODLY 2351
2281 KVFQGNQDSFTPVNSLDPPLTRYLRIHPQSWHVOIALRMEVLGCEAODLY 2332

RESULT 12
US-10-187-319-6
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 09/523,656
FILING DATE: 2000-03-10
APPLICATION NUMBER: US 09/037,601
FILING DATE: 1998-03-10
APPLICATION NUMBER: WO PCT/US97/11155
FILING DATE: 1997-06-26
APPLICATION NUMBER: US 08/670,707
FILING DATE: 1996-06-26
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
TITLE: Sequence of the Murine Factor VIII cDNA
JOURNAL: Genomics
VOLUME: 16

```

```

PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-187-319-6

Query Match
Best Local Similarity 72.1%; Score 8827; DB 14; Length 2319;
Matches 172; Conservative 232; Mismatches 364; Indels 54; Gaps 22;

QY 1 MQLIELTCCFLCLRLPCFSATRRYYIGAVELSDWQMOSD-LGELPVDARPPRPVPSKSPF 59
DB 1 MQLALFACPFLLSLFPCSSAIRRYVIGAVELSNWVQSDLLSVLHDSRFLPMSTSPF 60
QY 60 NTSVYVYKTKLFEVFDTHLFIENIAKPPPMGLLPTIQABVYTVVITLKNMASHVPSLHA 119
DB 61 NTSIMYKTKVFEYKQDLFNIAKPPPMGLLPTIWEVHDVTVITLKNMASHVPSLHA 120
QY 120 VGVSYWKASGAEYDDQTSOREKEDKVPFGSGSHYVWQVLKENGPMASDPLCLYSYLS 179
DB 121 VGVSYWKASGAEYDDQTSOREKEDKVPFGSGSHYVWQVLKENGPMASDPLCLYSYLS 180
QY 180 HVDLVKDLNSGLIGALLYVCREGLAKEKXTQLHKEFILLFAVDEGKSWHSETKNSLMQDR 239
DB 181 HVDLVKDLNSGLIGALLYVCREGLAKEKXTQLHKEFILLFAVDEGKSWHSETKNSLMQDR 240
QY 240 DAASARAPKPMHTVNGYVNRSLPGLIGCHRSYVYHVIHGMCTTPVHSIFLECHTFLVRN 299
DB 241 DSASARDAPKPMHTVNGYVNRSLPGLIGCHRSYVYHVIHGMCTTPVHSIFLECHTFLVRN 300
QY 300 HQASLEISPTITFLTAQTLMDLGOFLLSCHISSHQHDMGMEAYVVDSPCBEPQLRMK-N 358
DB 301 HQASLEISPTITFLTAQTLMDLGOFLLSCHISSHQHDMGMEAYVVDSPCBEPQLRMK-N 360
QY 359 NBEADYDDDDTDSMDVVRDDDDNSPSFIQIRSVAKKPKTWYHIAAEEDWDYAPLV 418
DB 361 NEEMEDYDDDDY-SEMDMFTLDYDSSP-FIQIRSVAKKPKTWYHIAAEEDWDYAPLV 418
QY 419 LAPDERSYKSYLNNGPORIGRKVKYKPMAYTDTFTKTRAIQHESGLPLLYGEVGD 478
DB 419 PTDNNGSYKSYLNNGPORIGRKVKYKPMAYTDTFTKTRAIQHESGLPLLYGEVGD 478
QY 479 TLLIIFKNOAGRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPLPGEIFKYKWTVYED 538
DB 479 TLLIIFKNOAGRPYNIYPHGITDVRPLYSRRLPKGVKHLKOPPLPGEIFKYKWTVYED 538
QY 539 GPTKSDPRCLTRYSSFFNMERDLASGLIGPLLYCYKESVDQGNQIMSDKRNVLFSVF 598
DB 539 GPTKSDPRCLTRYSSFFNMERDLASGLIGPLLYCYKESVDQGNQIMSDKRNVLFSVF 598
QY 599 DENRSWYLTENIQRFPLNPAGVQLEDPEFOASNMHSGINGVYFDSLOI-SVCLHEVAYWYI 658
DB 599 DENRSWYLTENIQRFPLNPAGVQLEDPEFOASNMHSGINGVYFDSLOI-SVCLHEVAYWYI 658
QY 659 LSVGAQTDPLSVFRSGYTFKHKMYVEDTLTLPFSGETVFMMSMENPGLIILGCHNSDPN 718
DB 659 LSVGAQTDPLSVFRSGYTFKHKMYVEDTLTLPFSGETVFMMSMENPGLIILGCHNSDPN 718
QY 719 RGMTHALLKVVSCDXTAGDYEDSDYEDISAYLLSKNAIEPRSPQNSRHPSTROKQFNAT 778
DB 719 RGMTHALLKVVSCDXTAGDYEDSDYEDISAYLLSKNAIEPRSPQNSRHPSTROKQFNAT 778
QY 779 TIPENDIEKTPDPAHRTMPKIONVSSDILLMLLROS-PTEPHGLSLDLOEAKYEFSD 837
DB 779 TIPENDIEKTPDPAHRTMPKIONVSSDILLMLLROS-PTEPHGLSLDLOEAKYEFSD 837
QY 838 DPSPCAIDSNNSLSEMTHTFRPQLHSGDMVFTSPSSQLRLNEXLGTATAELKCLDPKV 897
DB 839 DHSFNAIDSNNEGPKVQLRPESHSHSEKIVTTPFGQLRNSKSLFTIYVKKLGLQV 898
QY 898 SSTSNLLI-STIPSDNLAAGTNDTSSLPSPMPVHYDSQDITTLFGKKSSPLTSSGSPLS 956
DB 899 SSLPSNLMTTTTILSDNLKATPEKTDSSGFPDMPVHSSSKLSTTAFGKKAYSLVGSHPFN 958

```

QY 957 LSEENNDSKLLPSGLMNSOESSGKVSSTESGRFLFKGRAGCPALLTKONALFKVSI 1016
DB 959 ASZENSCLNLSLSTMYSESUPRDNILSIENDRLREKRFHGIALLTKONTLFDKNVSL 1018
QY 1017 LKTNKTSNNAIRKTHIDGPSLLENSPSVMQN--ILESDETFKVKVPLIHDRMLMDKNA 1075
DB 1019 MKTKTYNHTSTNEKLTESPT--SIENSTDLQDAILKYNSEIQEVYALIHGDTLAGKNS 1077
QY 1076 TALRLNHNSKNTTSSKNMEXVOQKGGPIPPDAQNPDMSFFKMLFPPESARWIORTHGN 1135
DB 1078 TYLRNLHNLNRITSTKNOIFHRKDEDPIDQEDNTIMPFSKMLFSESSENFKKTNGN 1137
QY 1136 SLNSGOCSPKQVLVSIGPEKSVGQNFLEKKNVVGKGEFTKDVGLKBMVPPSSNLF 1195
DB 1138 SLNSEGEHSPKQVLVLMFKKYKYNQSFLEKKNVVEQDGTKNIGSKDMAFPHNMSIFL 1197
QY 1196 TMLDNLHNNHTNQEKIIOEIEKKETLLOENVLPQIHVTGTGNPMKNLFLSTRQNV 1255
DB 1198 TILSNVHENGRENQEKNOEIEB--KEALTEEKVVLQVHEATGSKNFKLIDILITGRONI 1256
QY 1256 EGSYDGAAYAPVLODFPSLNDSTNRTKKHTAHSK--KGBEENLEGLGNQTKQIVEKYACT 1313
DB 1257 --SLYEVHVPVLQNTISNNSTNTVQIHMHEHFPKRAKDKETNSEGLVNTKREVMKVN--- 1311
QY 1314 TRISPNTSQNFVQTSKRALQKQRLPLEETELEKRIIVDDTSTOKSKMKHLLTPSTLTQ 1373
DB 1312 -----PSQKNITQTSKALGQRL-----STQWLKTINGSTQCIILQ 1349
QY 1374 IYNEKEKGAITOSPISDCLTSHSPQANRSPPIAKVSSPESIRPYLTVLRFVLFODNS 1433
DB 1350 IDHSEKMKKFIKSSLSDS--SVIKSTQTNSSDSHIVKTSAPP--PIDLKRSPFONKES 1405
QY 1434 HLPAAZY---RKDGSVOESHFQAGAKNLSLAILLEMTGQREVGSIGTSATNSV 1489
DB 1406 HVQASSYIYDFKTSRIQESNNFLKETKINNPSTAILPWNPIDGCKTSPCKSTNSV 1465
QY 1490 TYKKVENTVLPKDPKTSKGVKELLPKVHYQKDLFPETTSNGSPGHLDVSGSLQGTB 1549
DB 1466 TYKGRNIIFLAPTELESKGLLELPQVSIQBEILUPTTSHGSPHNLMEVFLOKIQ 1525
QY 1550 GAIKNEANRPGKVPFLRVATISSAKTPSKLDPLAWDNHYGQIPKBEWKQESKPEKT 1609
DB 1526 GPTKWKAKARHGE--SIKCKTSSKNTRSKLLNHAWDVHYAAQIPKDMWKSKEKSPBII 1583
QY 1610 AFKKKJTLISLACENSHAIANEQNKDELEVWAKQGRTERELCSQNPVVKRHRORI 1669
DB 1584 S-KQEDTILSLPHGNHSHSIGA--NEKQNPQRETTWVKQGTQRTCSQIPPVVKRHRORL 1642
QY 1670 TRTTIQQDQEEIDYDDTIIVENMKKEDFIYDDEQSPRSFQKTRHYFIAAVERLDYG 1729
DB 1643 --SAFQSEQEAITYDDAITIE--TIEDFDIYSEDIKQGRPSFOQKTRHYFIAAVERLDYG 1699
QY 1730 MSSSHVLNRAQSGSVPOFKVVPQFTDGTQPLRYGELNEHGLLGPYIRAEVDN 1789
DB 1700 NEST--HVLNRYQSDNVPOFKVVPQFTDGTQPLRYGELNEHGLLGPYIRAEVDN 1758
QY 1790 IAVTFNQAARSPVSYSSILSYEEDORQGAERKNEVKNETKYPWKQYHMAPTKQSF 1849
DB 1759 IAVTFNQAARSPVSYSSILSYEEDOR--GEBPRNFVKNETKIYPWKQYHMAPTKQSF 1817
QY 1850 DCKAWAYFSDVLEKDVHSLGIPLLIVCHTNTLNPARGSQVTVQEPALFTTFIDETKSWY 1909
DB 1818 DCKAWAYFSDVLEKDVHSLGIPLLI--CHANTLNPARGSQVTVQEPALFTTFIDETKSWY 1877
QY 1910 FTENVERNACACNIOEMDPTFKENTRFAINGYIMDTLPGLVMAQDQIRWYLLSMGN 1969
DB 1878 FTENVRKNCKTCFNFQEMDPTLKENYRFAINGYIMDTLPGLVMAQDQIRWYLLSMGN 1937
QY 1970 ENIHSHFSGHVTFTVKKEEYKVALNLYPGVPFVEMLPKSGATWRVSCILGEHLHAGM 2029
DB 1938 ENIQSHFSGHVTFTVKKEEYKQAVNLNLYPGVPFVEMLPKSGATWRVSCILGEHLHAGM 1997

QY 2030 STLELVYNNKCTPIGLMAGSHIRDFQITASGOYGQWAPKLARLHYSGSINAMSTKEPPSW 2089
DB 1998 STLELVYNNKCTPIGLMAGSHIRDFQITASGOYGQWAPKLARLHYSGSINAMSTKEPPSW 2057
QY 2090 IKVDLLAPMIHGIKTOGARQKFSLLYSQFIIMTSLDGKKWQTYRGNSTGTLMVFFGNV 2149
DB 2058 IKVDLLAPMIHGIKTOGARQKFSLLYSQFIIMTSLDGKKWLSYQGNSTGTLMVFFGNV 2117
QY 2150 DSSGIXHNIENPPIIARVIRLHPTHYSTRSTLMELMGCDLNSCMPGLMESKASDAOI 2209
DB 2118 DSSGIXHNSFNPIIARVIRLHPTHYSTRSTLMELMGCDLNSCIPGLMESKASDAOI 2177
QY 2210 TASSYFTNMFATWSPSKARLHLOGESNARPOVNNPKWLQVDFQXTMKVGTGVTGQVKS 2269
DB 2178 TASSYFTNMFATWSPSKARLHLOGESNARPOVNNPKWLQVDFQXTMKVGTGVTGQVKS 2237
QY 2270 LITSYVKEFLISSQDGHQWTLFTQNGKVKVQGNQDSFTFVNSLDPPLTRVLRHP 2329
DB 2238 LITSYVKEFLISSQDGHQWTLFTQNGKVKVQGNQDSFTFVNSLDPPLTRVLRHP 2297
QY 2330 QSWHQAIALRMEVLGCEAQQDY 2351
DB 2298 QIWEHQIALRLEILGCEAQQDY 2319

RESULT 13

US-10-131-510A-6
; Sequence 6, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 2319
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-131-510A-6

Query Match 71.1%; Score 8827; DB 14; Length 2319;
Best Local Similarity 72.5%; Ered. No. 0;
Matches 1712; Conservative 232; Mismatches 364; Indels 54; Gaps 22;

QY 1 MOELSTCFCLLFCFCSATRRYYLGAVELSWDYMQSD--LCELVDARFPPRPVKSPFF 59
DB 1 MQIALFAFFFLSLNFCSSAIRYYLGAVELSWYIQSDLLSVLFTDSRFLPRMSTSPFF 60
QY 60 NTSVYVYKTLFVFTDHLFNIAKPRPPWGLGPTIQAEVYDVTYITLKNMASHPVSLHA 119
DB 61 NTSIMYKKTVFVEYKQDLFNIAKPRPPWGLGPTIWTVEVHTVITLKNMASHPVSLHA 120
QY 120 VGVSVYKASEGAEYDQTSQREKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYSLYS 179
DB 121 VGVSVYKASEGDEYSDQTSQMEKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYSYNS 180

Db 2298 QIWEHQIALREILIGCRAQQOY 2319

RESULT 14

US-10-187-319-37

Sequence 37, Application US/10187319

Publication No. US20030068785A1

GENERAL INFORMATION:

APPLICANT: Lollar, John S.

TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

STREET: 5370 Manhattan Circle Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: USA

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/187,319

FILING DATE: 27-Aug-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/523,656

FILING DATE: 2000-03-10

APPLICATION NUMBER: US 09/037,601

FILING DATE: 1998-03-10

APPLICATION NUMBER: WO PCT/US97/11155

FILING DATE: 1997-06-26

APPLICATION NUMBER: US 08/670,707

FILING DATE: 1996-06-26

ATTORNEY/AGENT INFORMATION:

NAME: Greenlee, Lorraine L.

REGISTRATION NUMBER: 27,894

REFERENCE/DOCKET NUMBER: 75-95K

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303/499-8080

TELEFAX: 303/499-8089

INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:

LENGTH: 2133 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 37:

US-10-187-319-37

Query Match 65.0%; Score 8068; DB 14; Length 2133;

Best Local Similarity 67.0%; Pred. No. 0;

Matches 1577; Conservative 208; Mismatches 344; Indels 226; Gaps 17;

QY 1 MQELSTCFLLRFLCFCSATRRYYLGAVELSWDYMQSD-LGELPVDARFPFRVPSKPPF 59

Db 1 MQELSTCFLLRFLCFCSATRRYYLGAVELSWDYMQSD-LGELPVDARFPFRVPSKPPF 60

QY 60 NTSVYKTKLTFVETFDHLFNIAKPRPPWGLGPTIQAEVYDVTVITLKNMASHFVSLHA 119

Db 61 GPSVLYKTKTFVETFDQLFSVARPPWGLGPTIQAEVYDVTVITLKNMASHFVSLHA 120

QY 120 VGVSVMKASEAGYDDQTSOREKEDKVPFGSGSHYVMQVLKENGCPMADDPCLITYSYLS 179

Db 121 VGVSVMKASEAGYDDQTSOREKEDKVPFGSGSHYVMQVLKENGCPMADDPCLITYSYLS 180

QY 180 HVDLVKDLNSGLIGALLVCREGLAKEKTTQLTHKPFILFAVFDGKSWHSETQNSLMQDR 239

Db 181 HVDLVKDLNSGLIGALLVCREGLSRETQNLHNFVLLFAVFDGKSWHARNDSWTRAM 240

QY 240 DAASARAWFKKHTVNGYVNRSLPGLIGCHRSVYWEIVIGMGTTPVHSLFLEGTFLVRN 299

Db 241 DPAPARAQPAHHTVNGYVNRSLPGLIGCHRSVYWEIVIGMGTSPVHSLFLEGTFLVRH 300

QY 300 HROASLEISPIITLTACTLLMDLGQELLCHISSHOHGDNQEAUVKVDSCPEEOLRMKN 359

Db 301 HROASLEISPIITLTACTLLMDLGQELLCHISSHAHGMEAHVRVSCAESEPOLRRKAD 360

QY 360 EBAEDYDDDLTSEMVDVFRDDNSPSFIOIRSVAKKHPTKTVHYIAAEEEDMDYAPLVL 419

Db 361 EE-EDYDDNLYDSMDVRLDGDVSPFFIOIRSVAKKHPTKTVHYISABEDMDYAPAVP 419

QY 420 APDRSYKQYLNNGPQIRGRKYKVRFMAYTDETFTKREAIQHESGILGPLYGEVGT 479

Db 420 SPDRSYKSLYLNNGPQIRGRKYKVRFMAYTDETFTKREAIQHESGILGPLYGEVGT 479

QY 480 LLIIIFKNQASRPYNIYPHGITDVRLPYSLRPLPGVGHLDKDPILPGEIFKVKTVTVEDG 539

Db 480 LLIIIFKNQASRPYNIYPHGITDVRLPYSLRPLPGVGHLDKDPILPGEIFKVKTVTVEDG 539

QY 540 PTKSDPRCLTRYSSFVNBERDLASGLIGPLICYESVDQGNQIMSKRNVLFSVFD 599

Db 540 PTKSDPRCLTRYSSFVNBERDLASGLIGPLICYESVDQGNQIMSKRNVLFSVFD 599

QY 600 ENRSWYLTENIQRLPNPAGVQLEDPPFOASNMHESINGYVFDLSQLSVCLHEVAYWYL 659

Db 600 ENRSWYLTENIQRLPNPAGVQLEDPPFOASNMHESINGYVFDLSQLSVCLHEVAYWYL 659

QY 660 SIGAQTDFLSVFRSGYTFKHMYVEDTLTPPSGETVFMENPGLWILGCHNSDPRNR 719

Db 660 SIGAQTDFLSVFRSGYTFKHMYVEDTLTPPSGETVFMENPGLWILGCHNSDPRNR 719

QY 720 GMTALLKVSSCDKNTGYDESDYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATT 779

Db 720 GMTALLKVSSCDKNTGYDESDYEDISAYLLSKNNAIEPRSFQNSRHPSTROKQFNATT 779

QY 780 IPENDIEKTPWFAHRTPMKIQNVSSDDLMLLRQSPTHGLSLDLOAKYFTFSDDP 839

Db 780 IPENDIEKTPWFAHRTPMKIQNVSSDDLMLLRQSPTHGLSLDLOAKYFTFSDDP 839

QY 840 SPGAIDSNNSLSBMTHERPOLHSGDMVFTPESSQLRLNEKLGTATTELKLDKFKVSS 899

Db 840 SPGAIDSNNSLSBMTHERPOLHSGDMVFTPESSQLRLNEKLGTATTELKLDKFKVSS 899

QY 883 SSLLKTSPIIPSDTSAETHTSLGPPHPPQVNFRLGALVILGKNSHPIGAGVPLGS 942

Db 883 SSLLKTSPIIPSDTSAETHTSLGPPHPPQVNFRLGALVILGKNSHPIGAGVPLGS 942

QY 958 SEENNSKLLSEGLMNSQBSGKNSVSTSGRLFKGKRAHGPALLTKDNALFKVLSLL 1017

Db 958 SEENNSKLLSEGLMNSQBSGKNSVSTSGRLFKGKRAHGPALLTKDNALFKVLSLL 1017

QY 993 TEEDH-----ESSLGENVSPVSDGIFEKERAHGPASLTAKDDVLEKVNISLV 989

Db 993 TEEDH-----ESSLGENVSPVSDGIFEKERAHGPASLTAKDDVLEKVNISLV 989

QY 1018 KTNKTSNNSATNRKTHIDGPSLLTENSVPVQNLIESJTEPKVTPLIHDMMLDKNATA 1077

Db 1018 KTNKTSNNSATNRKTHIDGPSLLTENSVPVQNLIESJTEPKVTPLIHDMMLDKNATA 1077

QY 1078 LRLNEMNKTTSSKMEMVQCKKEGPIPPDAQNPDMSFPKMLFLPESARWIQTHGKNSL 1137

Db 1078 LRLNEMNKTTSSKMEMVQCKKEGPIPPDAQNPDMSFPKMLFLPESARWIQTHGKNSL 1137

QY 1029 SGLAHVN-----MIKGLGKNP 1047

Db 1029 SGLAHVN-----MIKGLGKNP 1047

QY 1138 NSGOGSPKQVLVSLGPKSVQGNFLSEKKNVUVKGGEFTKDVGLKEMVFPSSRNFLTN 1197

Db 1138 NSGOGSPKQVLVSLGPKSVQGNFLSEKKNVUVKGGEFTKDVGLKEMVFPSSRNFLTN 1197

QY 1048 SSERGPSPELLTSSSGSKVKGSSGQGRIRVAVVEEELSKG---KEMMLFNSLTLTN 1104

Db 1048 SSERGPSPELLTSSSGSKVKGSSGQGRIRVAVVEEELSKG---KEMMLFNSLTLTN 1104

QY 1198 LDNLHENHNTQEKIOEIEKTKETLIQENVLPIQHTVTGTKNFMKNLFLLSRQVVEG 1257

Db 1198 LDNLHENHNTQEKIOEIEKTKETLIQENVLPIQHTVTGTKNFMKNLFLLSRQVVEG 1257

QY 1105 SADVQGNHTSHQGGKSEEMERREKLVQEKVLDLPQVTAIGTKNPLNIFHQSTEPSVEG 1164

Db 1105 SADVQGNHTSHQGGKSEEMERREKLVQEKVLDLPQVTAIGTKNPLNIFHQSTEPSVEG 1164

QY 1258 SYDGAVAPVLQDFRSLNDS*NRKTKHTAHRPSKKEENLEGLNQCTQIVKACTTIRIS 1317

Db 1258 SYDGAVAPVLQDFRSLNDS*NRKTKHTAHRPSKKEENLEGLNQCTQIVKACTTIRIS 1317

QY 1165 FDGSHAPVQDSRLNDSAEARATHIAHPSAIREAPLEAPGNRT----- 1210

Db 1165 FDGSHAPVQDSRLNDSAEARATHIAHPSAIREAPLEAPGNRT----- 1210

QY 1318 PNTSQNFVTORSKRALQKRLPLEETELEKRIIVDDTISTQWSKMKHLTPSTLTQIDYN 1377

Db 1318 PNTSQNFVTORSKRALQKRLPLEETELEKRIIVDDTISTQWSKMKHLTPSTLTQIDYN 1377

```
Db 1211 -GPGPRSAVRVKQSLQIRLPLEEIKPGRGVNLNATSRWS----- 1252
Qy 1378 EKEGAIQTQSPSLDCLTRSHSIPZQANRSLPIAKVSSPPSIRPIYLTRVLFDQNSHLPA 1437
Db 1253 ----- 1252
Qy 1438 ASYRKXDSGVQSSSHFLOCAKNNLSLAILTLEMTGDQREVSLGTSAPNSVYTKVENT 1497
Db 1253 -----ESSPILOCAKNNLSLPFLTLEMAGGQKISALGKSAAGPLASGKLEKA 1301
Qy 1498 VLPKPDLPKTSKVELLPVHHYIYQKDLFTETSNOSPGHLDIVESLLOCTEGAIKNEA 1557
Db 1302 VLSSAGLSASAKAEFLPVVRHEDDLFEQKTSNVSCAHDGLGQIFLQKTRGPNVNLAK 1361
Qy 1558 NRPGKVPFLRVATSSAKTSPSKLLDPLAWDNHYGTQIPKEEMKSOEKSPKTAFAKKOTI 1617
Db 1362 NRPG-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALATKDI 1398
Qy 1618 -LSLNACESNHAIAINEQNKPEIEVWAKQRTERICSONPPVLKRHOREITRTTLOS 1676
Db 1399 SLPLDRHESNHSIAAKNEQAEAREAAWTKGGPRLCAPKPPVLLRRHQDISLPTFOP 1458
Qy 1677 DQEE-DYDDTIQVEMKEDFDIYDENOSPRSFOKTEHYFIAAVERLWDYCMSSPHV 1736
Db 1459 ESKMDYDDIFSTETKGEDFDIYGEDENQDPRSFOKTRHYFIAAVEQWLDYCMGESPPA 1518
Qy 1737 LNRQAQSGSVPOFKKVVFOEFTDGSFTOPLRYGELNEHLGLGPIYRAEVEDNIMVTFN 1796
Db 1519 LNRQAQNGVPRFKKVVREFADGSPQPSYRGELNKLGLGPIYRAEVEDNIMVTFN 1578
Qy 1797 QASRPSFYSSLSIIEVEDQGAEPKRNFKNEKTYFKVQHMHMAPTKDFDCKAWAY 1856
Db 1579 QASRPSFYSSLSIISYDDQGAEPHEFNVPQNETRTYFKVQHMHMAPTKDFDCKAWAY 1638
Qy 1857 FSDVLEKDVHSLIGLPLVCHTNTLPAHGRQVTVQEPALFTTIDETKSWFYFENMER 1916
Db 1639 FSDVLEKDVHSLIGLPLI CRANTLNAHGRQVTVQEPALFTTIDETKSWFYFENMER 1698
Qy 1917 NCRAPCNIQMEDPTFKENYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSMGSENHIS 1976
Db 1699 NCRAPCHQVEDPTLKENYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSMGSENHIS 1758
Qy 1977 FSGHYPTVRKKEEYKALVNLPGVPEETVEMLPKAGIWRVCECLIGEHLAGHMSILFLVY 2036
Db 1759 FSGHYPSVRKKEEYKAVNLPGVPEETVEMLPKAGIWRVCECLIGEHLAGHMSILFLVY 1818
Qy 2037 SNKQCTPLGMAQHIRDFOITASGOYQOWAPKLARLHYSGSINANSTKEPFWIKVDLLA 2096
Db 1819 SKECQAPLGMASGRIFDFOITASGOYQOWAPKLARLHYSGSINANSTKEPFWIKVDLLA 1878
Qy 2097 PMIHGKTQGARQKFPSSYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKH 2156
Db 1879 PMIHGIMTQGARQKFPSSYISQFIIMYSLDGRNMQSYRGNSTGTLMVFFGNVDSSGIKH 1938
Qy 2157 NIFNPPIARYFLRHPHYSIRSLRMLMGCDLNSCMPLGWESKAIASDAQITASSYFT 2216
Db 1939 NIFNPPIARYRLRHPHYSIRSLRMLMGCDLNSCMPLGWQKAIASDSQITASSHUS 1998
Qy 2217 NMFATWSPKARLHLQGRNAMEPQVNNPKMLQVDFQTKMKTGTGTGQVKSILITSMVY 2276
Db 1999 NIFATWSPQARLHLQGRNAMEPRVSSAEMQLVDLQTKVKTGTGTGQVKSILSSMYV 2058
Qy 2277 KEPLISSQDGHOWITLFFQNGKVKVFGQNDSTFPVNSLDPLPLTRYLRHPQSWHIOI 2336
Db 2059 KEPLVSSQDGRWITLFLQDGHTKVFGQNDSTFPVNSLDPLPLTRYLRHPQSWHIOI 2118
Qy 2337 ALRMEVLGCEAODLY 2351
Db 2119 ALRLEVLGCEAODLY 2133
```

RESULT 15

US-10-131-510A-37

```
; Sequence 37, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131.510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2133
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-131-510A-37
```

```
Query Match 65.0%; Score 8068; DB 14; Length 2133;
Best Local Similarity 67.0%; Pred. No. 0;
Matches 1577; Conservative 208; Mismatches 344; Indels 236; Gaps 17;

Qy 1 MOELSTCFECLLAFCSATRRYVLGAVLSWDVMSQSD-LGELPVDARFEPVRPKSPFP 59
Db 1 MOELSTCVFCLLPGLGFSARRYVLGAVLSWDVRSQSELRRELVHVDTRFPATAPCALPL 60

Qy 60 NTSVYKTLFVEFTDHLFNIAKPRPPPMWGLLGPITQAEVYDVVITLKNWASHPVSLHA 119
Db 61 GFSVLYKKTVPVEFTDQLFSVARPPPMWGLLGPITQAEVYDVVITLKNWASHPVSLHA 120

Qy 120 VCVSYWKASGAEYDDQTSQREKEDDKVPFGSHTYVWQVLKENGPMASDPLCLITYSL 179
Db 121 VCVSWKSEGAEYSDHTSQREKEDDKVLPGKSQTYVWQVLKENGPTASDPPCLITYSL 180

Qy 180 HVDLVKDLNSGLIGALLVCREGLAKETQTLHKETILLPAVFDEGKSWHSETKNSLMODR 239
Db 181 HVDLVKDLNSGLIGALLVCREGLTRETQNLHEFVLIPAVFDEGKSWHSAENDSWTRAM 240

Qy 240 DAASARAMPKMTVNGYVNRSLPGLIGCHRKSVYVHVIGMGTTPVHGFIFLEGHTFLVRN 299
Db 241 DPAPARAQAPAMHTVNGYVNRSLPGLIGCHKSVYVHVIGMGTSPVHGFIFLEGHTFLVRH 300

Qy 300 HQASLEISPTIFLTAOTLLMDLGQFLSCHLSSHQHGMAYKVVDSCPRPQLRMKNX 359
Db 301 HQASLEISPTIFLTAOTLLMDLGQFLSCHLSSHHGMEAHVESCAEPQLRRKAD 360

Qy 360 EBAEDYDDDLTDSEMDVVRFDNDNSPFTQIRSVAKKPKTWVHYIAAEEDWDYAPLVL 419
Db 361 EB-EDYDNLVSDMDVVRDLGDDVSPFIQIRSVAKKPKTWVHYISAEEDWDYAPAVP 419

Qy 420 APDSRSYKQYVNNQPGIRGRYKVRMAYTDETFKTRTREAIOHESGILGPLLYGEVGT 479
Db 420 SPDSRSYKSLVNSGPGIRGRYKVRMAYTDETFKTRTREAIOHESGILGPLLYGEVGT 479

Qy 480 LLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFKVKWTVTVSDG 539
Db 480 LLIIFKNKASRPYNIYPHGITDVSALHPCRLLKGVKHLKDPILPGEIFKVKWTVTVSDG 539

Qy 540 PTKSDPRCLTRYSSSFVNMERDLASGLIGPLLIICYKESVDQGRNQIMSDKRNVLFSVFD 599
Db 540 PTKSDPRCLTRYSSSINLEKDLASGLIGPLLIICYKESVDQGRNQIMSDKRNVLFSVFD 599
```

QY 600 ENSGWJTEIQRFLPNPAGVQLEDEPEFOASNIMESINGVYFDSLOLSVCLHEVAYWYL 659
Db 600 ENSGWLAENIQFLPNPDLGLQODPEFOASNIMHSINGVYFDSLOLSVCLHEVAYWYL 659
QY 660 SIGAQTDFLSPSGYTFKHKMYEDTLTLFPSPGETVFMNMPGLWLGCHNSFRNR 719
Db 660 SVGAQTDFLSPSGYTFKHKMYEDTLTLFPSPGETVFMNMPGLWLGCHNSDLNR 719
QY 720 GMTALLKVVSCDKNITGDYEDSYEDISAYLLSKNNAIEPSSONSRHHPSTROKOFNAT 779
Db 720 GMTALLKVYSCDRDIDGYDNTVEDIPGLLSKNVIEPRSPQNRPPSASOKOQIT 779
QY 780 IPNDIEKTPFAHPTMPKIQNVSSDILLMLLROSPTPHGLSLDLOBAKYETFSDDP 839
Db 780 SPEDDV3-LDPOSGERTQALEELSVPSGDSMLLGNPAPHGSSSDLOQBARNE--ADDY 836
QY 840 SPGAIDSNLSSEWTHRPOLHSGDWVFTPESGLQRLNKEKLTAAATELKXLDKPVSS 899
Db 837 LPGARERNTAPSAARLRPBLHSAERVITPEP-----EK-----ELKKLDSKMS 882
QY 900 TSNMLIS--TIPSDNLAAGDNTSSLGPPSPVHYDSQDITTLFGKKSSPLTBSGGLSL 957
Db 883 SSDLLKTSPTIPSDTISAETHERTHSGPHQVNFPSQICAIVLGKNSHFICAGVPLGS 942
QY 958 SEENDSKULLESGLMNSQSSWGKANSYTESGBLFGKRAHGPAALTKONALFKVSI 1017
Db 943 TEEDH-----BSSLGENVSPVESDGIFEKERAHGASLTXXDVLFKVNISLV 999
QY 1018 KYNKTSNNSATNRKTHIDGPSILLIENSPSVQWQILSDTEFKKVTPLIHDRMLDKNATA 1077
Db 990 KYNKARVYLKTRKHIDDAALLTENRASA-----TFMDXNTTA 1028
QY 1078 LRLNHSNKTTSKKNMEMVQKKEGPIPPDAQNPDMSFFMFLPSSARWIKORTGKNSL 1137
Db 1029 SGLNHVN-----WIKPLGKNFL 1047
QY 1138 NSGQSPKQVLSLGEKVEHQNFUSEKKNVVGKGETKDVGLKBMVPPSSRNLF 1197
Db 1048 SSERGPSPELLTSSSGSGKVKQSSQGRIRVAVEBELSKG---KEMMLPNSLFTLN 1104
QY 1198 LBNLHNNTHNOKKIQRIEKKETLQIENVVLPOHTVTGKFMKNUFLISTRONVEG 1257
Db 1105 SADVQNDTHSGOKKREMERREKLVQBKVDLPQVYTAIGTKFLNFIHQSTEPSVEG 1164
QY 1258 SYDGYAPVLQDFRSLNDSNRKXTHAFSKKGBEENLEGKNTKQIVEXYACTTRIS 1317
Db 1165 FDGGSHAPVQDSRLNDSARAETHIAHFSARSEAPLEAPGNRI-----TFMDXNTTA 1210
QY 1318 PNTSQNFVORSKALKQFRIPLBETELEKRIIIVDDTSTQWSKNKHLPSTLTQIDYN 1377
Db 1211 -GPGPRSAVPRVKOSLQIRLPLEBEIKPDERGVNLNATSRWS----- 1252
QY 1378 EKEKGAITQSPSLDCLTRSHSIPQANRSLPIAKVSSPSPSIRPIVILTRVLFQDNSSHLPA 1437
Db 1253 ----- 1252
QY 1438 ASYRKDSCGVQSSHFLOGAKANNLSAILTLEMTGDQREVGLSGTSATNSVYKKVENT 1497
Db 1253 -----ESSPILQCAKNNLSPLTLLEMAGGOKISALGKSAAGPLASCKLEKA 1301
QY 1498 VLPKPDLPKTSKVELLPKVHIYQKDLPEPTESNCSPOHLDLVEGSLIQGTGAIKWBNA 1557
Db 1302 VLSAGLSEASGKAEFLPKVRVHRBDLLPQKTSNVSCAHGDJGQBIFLQKTRGPVNLNKV 1361
QY 1558 NRPKGVPFLRVATSSAKTPSKLLPLAWDNHYGTQIPKEWKSOKSPKTAFAKKDTI 1617
Db 1362 NEPG-----RTPSKLGP-----PMFK-EWESLEKSPKSTALTKDII 1398
QY 1618 -LSLNACSNHAIAAINEGONKPEIEVTWAKGRTERLCSONPPVLKHEQREITRTTLQS 1676
Db 1399 SLPLDRHSNHSIAAKNEGQAEQAEQAAWTKGGGRLCAPKPPVLRHRHQDISLPTQP 1458

QY 1677 DOBEIDYDDTISVEMKKEDFDIYDEBENQSPRSFOKTEHYFTAAVERLWDYCMSSPHV 1736
Db 1459 EEDXMDYDIFSTETKGEDFDIYGEDENQDPRSFOKTRHYFTAAVEQWDYCMSSPRA 1518
QY 1737 LRNRAGSGSVPOFKKVVFOEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFNR 1796
Db 1519 LRNRAGNGEVPRFKKVVREFADGSGFTQSYRGELNKLGLLGPYIRAEVEDNIMVTFNR 1578
QY 1797 QASRPYSFSSILISYEDORQAEPRKNVKNENKTYFKVQVHMAPTKDBPDCCKAWAY 1856
Db 1579 QASRPYSFSSILISYDDEQAEPRHNFVQPNETRTYFKVQVHMAPTDEBDCCKAWAY 1638
QY 1857 FSDVDLEKDVHSLIGLPLVCHNTLNPAHGQVTVQEFALPRTTIFDEKSWYFTNMER 1916
Db 1639 FSDVDLEKDVHSLIGLPLLCRANTLNAAHGQVTVQEFALPRTTIFDEKSWYFTNMER 1698
QY 1917 NCRAPONIOMEDPTFKENYRPHALNGYIMDTLPLGVAQDQIRWYLLSGNSNENIHSIH 1976
Db 1699 NCRAPCHLQMEDPTLKENYRPHALNGYVMDTLPGLVMAQNRIRWYLLSGNSNENIHSIH 1758
QY 1977 FSGHYFTVKKBEYKXALYNLYPGVPETVEMLPKSRAGIWRVCLIGEHLHAGNSTLFLVY 2036
Db 1759 FSGHYFTVKKBEYKXAVNLYPGVPETVEMLPKSVGIWRIECLIGEHLQAGNSTLFLVY 1818
QY 2037 SNKCCPTPLGMAHGHRDFOITASGOVOWAPKLARLHYSGSINAKSTKEPPSKIKVDLLA 2096
Db 1819 SKXCQAPLGMAHGRINDFOITASGOVOWAPKLARLHYSGSINAKSTKDPHSWIKVDLLA 1878
QY 2097 PMIHGIMTQGARQKPFSSLYISQFIIMYSLDGKKQWQYRGNSTGTLMVFPVGNVDSGK 2156
Db 1879 PMIHGIMTQGARQKPFSSLYISQFIIMYSLDGRNWSYRGNSTGTLMVFPVGNVDSGK 1938
QY 2157 NIFNPPIIARYIRLHPHTHYSTIRLMELMGCDLNSCMLPGLWESKASIDAOITASSYFT 2216
Db 1939 NIFNPPIIARYIRLHPHTHYSTIRLMELMGCDLNSCMLPGLWESKASIDAOITASSYFT 1998
QY 2217 NMFATWSPSKARLHLQGRSNARPOVNNPKELQVDFOKTMKVTVTTQGVKSLTSMV 2276
Db 1999 NIFATWSPSKARLHLQGRSNARPOVNNPKELQVDFOKTMKVTVTTQGVKSLTSMV 2058
QY 2277 KEFLISSQDGHQWTLFPQNGKVKVFGNQDSTFPVANSPLDPLLRYLIRIHPOSVWQHI 2336
Db 2059 KEFLVSSQDGRWTLFLQDGHKVKVFGNQDSTFPVANSPLDPLLRYLIRIHPOSVWQHI 2118
QY 2337 ALRMEVLGCEADLY 2351
Db 2119 ALRMEVLGCEADLY 2133

Search completed: April 13, 2004, 14:19:52
Job time : 73 secs

XX SQ Sequence 2351 AA;
 Query Match 99.9%; Score 12399; DB 1; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQIELSTCFECLLRFCFSATRYVYLGAVELSWDYMSDGLGELPVDARPPRPVVKPFPPN 60
 DB 1 MQIELSTCFECLLRFCFSATRYVYLGAVELSWDYMSDGLGELPVDARPPRPVVKPFPPN 60

QY 61 TSVVYKTLFVETTDHLEN-AKRRPPPMWGLLGPITQAEVYDVTVITLKNWASHPVSLHAV 120
 DB 61 TSVVYKTLFVETTDHLEN-AKRRPPPMWGLLGPITQAEVYDVTVITLKNWASHPVSLHAV 120

QY 121 GSVYWKASGAEGVDDQTSOREKDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSYLSH 190
 DB 121 GSVYWKASGAEGVDDQTSOREKDDKVPFGSGSHYVQVLKENGPMASDPLCLTYSYLSH 190

QY 181 VDLVKOLNSGLIGALLVCREGSLAKEXTQTLHRKFIILLFAVFDGKSWHSETKNSLMQDRD 240
 DB 181 VDLVKOLNSGLIGALLVCREGSLAKEXTQTLHRKFIILLFAVFDGKSWHSETKNSLMQDRD 240

QY 241 AASARAWPKMHTVNGYVWRSIPGLIGCHRKSVMTHVIGMGTTPGVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPKMHTVNGYVWRSIPGLIGCHRKSVMTHVIGMGTTPGVHSIFLEGHTFLVRNH 300

QY 301 RQASLEISPIITFTAQTLMDLQGLFLLSCHISSHQHDMGEAYVKVDSCPEEPOLIMKNNE 360
 DB 301 RQASLEISPIITFTAQTLMDLQGLFLLSCHISSHQHDMGEAYVKVDSCPEEPOLIMKNNE 360

QY 361 EADYDDDLTDSMDVVRFDNDDSPFIQIRSVAKKHPKTHVHYIAAEEDWDYAPLVLA 420
 DB 361 EADYDDDLTDSMDVVRFDNDDSPFIQIRSVAKKHPKTHVHYIAAEEDWDYAPLVLA 420

QY 421 PDERSYKSOVLNNGPQRIGRKYKVRPMAYTDFTFKTREAIQHESGILGPLLYGEGVDTL 480
 DB 421 PDERSYKSOVLNNGPQRIGRKYKVRPMAYTDFTFKTREAIQHESGILGPLLYGEGVDTL 480

QY 481 LIIFKQASRPYNIYPHGITDVRPLSRRLPKGVKHLKDPFPLPGRIFKYKMTVTVEDGP 540
 DB 481 LIIFKQASRPYNIYPHGITDVRPLSRRLPKGVKHLKDPFPLPGRIFKYKMTVTVEDGP 540

QY 541 TKSDPCLTRYSSFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSDKRNVLFSVFDE 600
 DB 541 TKSDPCLTRYSSFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSDKRNVLFSVFDE 600

QY 601 NRSWYLTENIQRLPNPAGVOLBDPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILS 660
 DB 601 NRSWYLTENIQRLPNPAGVOLBDPFOASNIHMSINGYVFDLSQLSVCLHEVAYWYILS 660

QY 661 IGAQTDPLSVFFSGYTFKHVMVYEDTLTLPFFSGETVFMSENPGLWILGCHNSDFRNRG 720
 DB 661 IGAQTDPLSVFFSGYTFKHVMVYEDTLTLPFFSGETVFMSENPGLWILGCHNSDFRNRG 720

QY 721 WTALLKVSXCDKNTGYIDSYBDISAYLLSKNAIEPRSFNSRHPSTROKQFNATTI 780
 DB 721 WTALLKVSXCDKNTGYIDSYBDISAYLLSKNAIEPRSFNSRHPSTROKQFNATTI 780

QY 781 PENDIEKTDPPFAHRTPMPKIQNVSSDDLMLLRQSPTEPHGLSLDQAKYETPSDDPS 840
 DB 781 PENDIEKTDPPFAHRTPMPKIQNVSSDDLMLLRQSPTEPHGLSLDQAKYETPSDDPS 840

QY 841 PGAIIDNNLSLSETHFRPQLHSGDMVFTPESGQLRLNEKLGTTAATLKLKLDKFKVST 900
 DB 841 PGAIIDNNLSLSETHFRPQLHSGDMVFTPESGQLRLNEKLGTTAATLKLKLDKFKVST 900

QY 901 SNNLJSTIPSDNLAAGTDNTSSLGPPSMVHYDSDQDITLFGKKSPLTESGGLSLSEE 960
 DB 901 SNNLJSTIPSDNLAAGTDNTSSLGPPSMVHYDSDQDITLFGKKSPLTESGGLSLSEE 960

QY 961 NNDKLLSEGLMNSQBSWGNVSVSTESGRLFKGKAHGPALLTKDNALFKVYSISLLKTN 1020
 DB 961 NNDKLLSEGLMNSQBSWGNVSVSTESGRLFKGKAHGPALLTKDNALFKVYSISLLKTN 1020

DB 961 NNDKLLSEGLMNSQBSWGNVSVSTESGRLFKGKAHGPALLTKDNALFKVYSISLLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNIILBSDTFFKKVTPLIHDMMLMDKNAALRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNIILBSDTFFKKVTPLIHDMMLMDKNAALRL 1080

QY 1081 NEWSNKTTSKKNEMVQOKKEGPIPPDAONPDMSPFKMLPLPESAEWIORTEGKNSLNG 1140
 DB 1081 NEWSNKTTSKKNEMVQOKKEGPIPPDAONPDMSPFKMLPLPESAEWIORTEGKNSLNG 1140

QY 1141 QQSPKQOLVSLGPEKSEVEGQNFISEKNKVVGKGEFTKDVGLKEMVFPSPSRNLFNLN 1200
 DB 1141 QQSPKQOLVSLGPEKSEVEGQNFISEKNKVVGKGEFTKDVGLKEMVFPSPSRNLFNLN 1200

QY 1201 LHENNTNHOEKIKQEBIEKKEKTLIQENVVLPOIHTVTGTFKNFMKNLFLSTEQNVESYD 1260
 DB 1201 LHENNTNHOEKIKQEBIEKKEKTLIQENVVLPOIHTVTGTFKNFMKNLFLSTEQNVESYD 1260

QY 1261 GAYAPVLODFRSLNDSTNRTKHTAHFSKKGSEENLEGLGNOTKOIVEKYACTRISPN 1320
 DB 1261 GAYAPVLODFRSLNDSTNRTKHTAHFSKKGSEENLEGLGNOTKOIVEKYACTRISPN 1320

QY 1321 SQONFVTSQSKALQKFRPLBETELEKIIIVDDTSTOWSKMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SQONFVTSQSKALQKFRPLBETELEKIIIVDDTSTOWSKMKHLTPSTLTQIDYNEKE 1380

QY 1381 KGAITQSPUSDCLTRSHSIPQANRSPPLPIAKVSSPSPRIPIYLTAVLPDONSHPAAS 1440
 DB 1381 KGAITQSPUSDCLTRSHSIPQANRSPPLPIAKVSSPSPRIPIYLTAVLPDONSHPAAS 1440

QY 1441 RKXDSGVQBSHFLQAKKNLSLAILTLEMTGDQREVSLGTSATNSVYKAKVENTVLP 1500
 DB 1441 RKXDSGVQBSHFLQAKKNLSLAILTLEMTGDQREVSLGTSATNSVYKAKVENTVLP 1500

QY 1501 XPDLPKTSQKVELLPKVHIYQKDLPTTETNSGSPGHLDIVBGLSLQGTGAKWNEANRP 1560
 DB 1501 XPDLPKTSQKVELLPKVHIYQKDLPTTETNSGSPGHLDIVBGLSLQGTGAKWNEANRP 1560

QY 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGQIPKSEKSKOEKPEKTAFAKKKTIISL 1620
 DB 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGQIPKSEKSKOEKPEKTAFAKKKTIISL 1620

QY 1621 NACESNHAIAALINEGONKPEIEVTWAKQGRTERLCSQNPVVKRQREBITRITLOSDOB 1680
 DB 1621 NACESNHAIAALINEGONKPEIEVTWAKQGRTERLCSQNPVVKRQREBITRITLOSDOB 1680

QY 1681 IDYDDTIISVEMKKEPDIYDEBENQSPRSFQKTRHYTAAVERLDWYGMSSSPHVRNR 1740
 DB 1681 IDYDDTIISVEMKKEPDIYDEBENQSPRSFQKTRHYTAAVERLDWYGMSSSPHVRNR 1740

QY 1741 AQSGVPOPKVVOFDTGDSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFEAOASR 1800
 DB 1741 AQSGVPOPKVVOFDTGDSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFEAOASR 1800

QY 1801 PYSPYSSLIISYBEDQOQAEPRKNTVKPNETKTYFWKVQHMAPTKDBDFCKAWAYFSDV 1860
 DB 1801 PYSPYSSLIISYBEDQOQAEPRKNTVKPNETKTYFWKVQHMAPTKDBDFCKAWAYFSDV 1860

QY 1861 DLEKDVHSLGLPLAVCHTNTLNPAHGRQVTVQBPALFTTIDETKSWYFTENMERNCRA 1920
 DB 1861 DLEKDVHSLGLPLAVCHTNTLNPAHGRQVTVQBPALFTTIDETKSWYFTENMERNCRA 1920

QY 1921 PCNIOEMDPTFKENYFRHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGH 1980
 DB 1921 PCNIOEMDPTFKENYFRHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGH 1980

QY 1981 VFTVAKKEEYKVALNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
 DB 1981 VFTVAKKEEYKVALNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040

QY 2041 QTPPLGWSGHIRDFOITASGQVQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
 DB 2041 QTPPLGWSGHIRDFOITASGQVQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100

QY 2101 HGKIQGARQKSSLYISQFIIMYSLDGKKWQTYGNSGTGLMVFFGNVDSSGIKHNIFN 2160
 DB 2101 HGKIQGARQKSSLYISQFIIMYSLDGKKWQTYGNSGTGLMVFFGNVDSSGIKHNIFN 2160
 QY 2161 PPIIARYIRLHPHTYSIRSLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMYFA 2220
 DB 2161 PPIIARYIRLHPHTYSIRSLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMYFA 2220
 QY 2221 TWSPSKARLHLQGRNARPOVNNPKWLOVDFOKTMKVGTGTVGKSLTSMYVKFEL 2280
 DB 2221 TWSPSKARLHLQGRNARPOVNNPKWLOVDFOKTMKVGTGTVGKSLTSMYVKFEL 2280
 QY 2281 ISSSQDGHWTLPFGNGKVKVQGNQDSFTPVVNSLDPELLTRYLRIRHPOSMWHQIALRM 2340
 DB 2281 ISSSQDGHWTLPFGNGKVKVQGNQDSFTPVVNSLDPELLTRYLRIRHPOSMWHQIALRM 2340
 QY 2341 EVLGCEAODLY 2351
 DB 2341 EVLGCEAODLY 2351

RESULT 2

AAW11329
 ID AAW11329 standard; protein; 2351 AA.

AC AAW11329;

DT 17-NOV-1997 (first entry)

XX DE Native Factor VIII:C.

XX KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophilic;
 KW proteolytic cleavage.

OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1...19

FT Protein /note= "signal peptide"

FT Region /note= "mature Factor VIII:C"

FT Domain /note= "heavy chain fragment"

FT Region /note= "B domain"

FT Region /note= "ligand chain fragment"

XX WO9703195-A1.

XX 30-JAN-1997.

XX 09-JUL-1996; 96WC-US011444.

XX 11-JUL-1995; 95US-0001025P.

XX (CHIR) CHIRON CORP.

XX Hung DT, Cohen FE, Innis M;

XX WPI; 1997-119050/11.

XX N-PSDB; AAT51357.

XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -

XX used in the treatment of haemophilias, by improvement of haemostasis.

XX Disclosure; Fig 1; 90pp; English.

XX This sequence represents the native Factor VIII:C. Factor VIII:C is a

XX large glycoprotein that participates in the blood coagulation cascade

CC that ultimately converts soluble fibrinogen to insoluble fibrin clot,
 CC effecting haemostasis. A deficiency in Factor VIII:C is responsible for
 CC haemophilia A, which is an X-chromosome-linked inherited bleeding
 CC diathesis. Factor VIII:C is activated by plasma proteases, such as
 CC thrombin. During activation the mature polypeptide is cleaved to generate
 CC heavy and light chain fragments that are further cleaved. The DNA
 CC encoding this sequence is mutated, using mutagenic primers, to produce
 CC the active Factor VIII:C analogues of the invention (such as AAW11330).
 CC The analogues comprise a native Factor VIII:C polypeptide modified at a
 CC site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-
 CC Arg dipeptide is created. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 XX SQ Sequence 2351 AA;

Query Match 99.9%; Score 12399; DB 2; Length 2351;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MQEILSTCFELCLLRFCSATRRYYLGAVERISWDYMQSDLGELPVDARPPRVPKSFN 60

DB 1 MQEILSTCFELCLLRFCSATRRYYLGAVERISWDYMQSDLGELPVDARPPRVPKSFN 60

QY 61 TSVVYKTLFVEFTDHLFNIAKPPPPMGLLGPITQAEVDTVTITLKNMASHPVLSHAV 120

DB 61 TSVVYKTLFVEFTDHLFNIAKPPPPMGLLGPITQAEVDTVTITLKNMASHPVLSHAV 120

QY 121 GVSYYKASEGAEYDDQTSQREKDDKVPGGSHYTVWQVLKENGSMADPCLTYSYLSH 180

DB 121 GVSYYKASEGAEYDDQTSQREKDDKVPGGSHYTVWQVLKENGSMADPCLTYSYLSH 180

QY 181 VDLVKDNLNSLIGALLVCREGSLAKEKTTQTLHKFILLFAVDEGKSHSETKNSLMQDRD 240

DB 181 VDLVKDNLNSLIGALLVCREGSLAKEKTTQTLHKFILLFAVDEGKSHSETKNSLMQDRD 240

QY 241 AASRAWPKMHTVRGYVNRSLPGLIGCHRKSVYWEHVGIMGTTPVHSIFLEGHTFLVRNH 300

DB 241 AASRAWPKMHTVRGYVNRSLPGLIGCHRKSVYWEHVGIMGTTPVHSIFLEGHTFLVRNH 300

QY 301 QASLEISPIITFLTAQTLMDLQGLI SCHTSSHQHDCGAEYKVDSCPEEPQLAKKNE 360

DB 301 QASLEISPIITFLTAQTLMDLQGLI SCHTSSHQHDCGAEYKVDSCPEEPQLAKKNE 360

QY 361 EAEDYDDDLTDSMDVVRFDNDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420

DB 361 EAEDYDDDLTDSMDVVRFDNDNSPSFIQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420

QY 421 PDRSYKQYLNNGPQIRGRYKVRPMAYTDETFKTREAI CHESGILGLLYGEVGTDL 480

DB 421 PDRSYKQYLNNGPQIRGRYKVRPMAYTDETFKTREAI CHESGILGLLYGEVGTDL 480

QY 481 LLIIFKNQASRPVNTYPHGITDVRPLYSERLPKGVKHLKDPILPGEIIFKYKWTVTVEDGP 540

DB 481 LLIIFKNQASRPVNTYPHGITDVRPLYSERLPKGVKHLKDPILPGEIIFKYKWTVTVEDGP 540

QY 541 TKSDPRCLTRYSSFFVNMERDLASGLIGPLLI CYKESVDQGNQIMSDKRVNVLFSVFDE 600

DB 541 TKSDPRCLTRYSSFFVNMERDLASGLIGPLLI CYKESVDQGNQIMSDKRVNVLFSVFDE 600

QY 601 NRSWYLTENIQRFLPNPAGVQLEDPEFOASNMHMSINGYVFDLSQLSVCLHEVAYWYLS 660

DB 601 NRSWYLTENIQRFLPNPAGVQLEDPEFOASNMHMSINGYVFDLSQLSVCLHEVAYWYLS 660

QY 661 IGAQTDFLSVFSGYTFKHKVVYEDTLTLPFSGETVFMSENPCILWLGCHNSDFNRG 720

DB 661 IGAQTDFLSVFSGYTFKHKVVYEDTLTLPFSGETVFMSENPCILWLGCHNSDFNRG 720

QY 721 MTALLKVVSSCDXNTGYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFVATTI 780
DB 721 MTALLKVVSSCDXNTGYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTROKQFVATTI 780
QY 781 PENDIEKIDPFAHRTMPKIQNVSSDILLMLLRQSPHPHGLSLDLQAKVETSDDDPS 840
DB 781 PENDIEKIDPFAHRTMPKIQNVSSDILLMLLRQSPHPHGLSLDLQAKVETSDDDPS 840
QY 841 PGAIIDNNLSSEMTFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDPKVSST 900
DB 841 PGAIIDNNLSSEMTFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATLKKLDPKVSST 900
QY 901 SNNLISTIPSDNLAAGTNDTSSLGPPSPVHYDSQDITTLFGKKSSPLTSGGPLSSEE 960
DB 901 SNNLISTIPSDNLAAGTNDTSSLGPPSPVHYDSQDITTLFGKKSSPLTSGGPLSSEE 960
QY 961 NNDKSLIEGLMNSQSSWGKNVSSTESGELFKGKAHGPALLTKDNALPKVSIISLLKTN 1020
DB 961 NNDKSLIEGLMNSQSSWGKNVSSTESGELFKGKAHGPALLTKDNALPKVSIISLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSPVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSPVWQNI LESDTEFKKVTPLIHDRMLMDKNATALRL 1080
QY 1081 NMSNKTTSKNMVMYQKKEGIPDPAQNPDMSFFKMLFLPESARWIQRTGKNSLNSG 1140
DB 1081 NMSNKTTSKNMVMYQKKEGIPDPAQNPDMSFFKMLFLPESARWIQRTGKNSLNSG 1140
QY 1141 QGSPKQVLSLGPKEKVEGQNFISEKNKVVGGEETKQGLKEMVFPSSRNILFTNLN 1200
DB 1141 QGSPKQVLSLGPKEKVEGQNFISEKNKVVGGEETKQGLKEMVFPSSRNILFTNLN 1200
QY 1201 LHENNTNHEKKTQEBIEKKEKTELIQENVLPQIHTVTGKNFMKNI FLASTRONVEGSD 1260
DB 1201 LHENNTNHEKKTQEBIEKKEKTELIQENVLPQIHTVTGKNFMKNI FLASTRONVEGSD 1260
QY 1261 GAYAPVLQDPRSLNDSTNRKTAHFSKKEBENLEGLNQTKQIIVEKYACTTRISPNT 1320
DB 1261 GAYAPVLQDPRSLNDSTNRKTAHFSKKEBENLEGLNQTKQIIVEKYACTTRISPNT 1320
QY 1321 SQQNFVQTSKRALKOPRLPLEBTELEKRIIVDDTSTQMSKNKMLTPSTLTQIDYNEKE 1380
DB 1321 SQQNFVQTSKRALKOPRLPLEBTELEKRIIVDDTSTQMSKNKMLTPSTLTQIDYNEKE 1380
QY 1381 KGAIQSPSLSDCJTRGSHIPQANRSELPQAKVSSFPSIRPIYITRVLFDQNSSHLPAASY 1440
DB 1381 KGAIQSPSLSDCJTRGSHIPQANRSELPQAKVSSFPSIRPIYITRVLFDQNSSHLPAASY 1440
QY 1441 RKDQSGVQSSSHFLOGAKKXNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
DB 1441 RKDQSGVQSSSHFLOGAKKXNLSLAILTLEMTGDQREVGLGTSATNSVTYKKVENTVLP 1500
QY 1501 KPDLPTSGKVLLPKVHIYOKDLFPPTETNSGSPGHLDLVEGSLLOQTEGAIKWNANRP 1560
DB 1501 KPDLPTSGKVLLPKVHIYOKDLFPPTETNSGSPGHLDLVEGSLLOQTEGAIKWNANRP 1560
QY 1561 GKVPFLVATESSAKTSPSKLLDPLAWDNHYGTQIPKEEMKSQBSPEKTAFAKKDKTILSL 1620
DB 1561 GKVPFLVATESSAKTSPSKLLDPLAWDNHYGTQIPKEEMKSQBSPEKTAFAKKDKTILSL 1620
QY 1621 NACESNHAIAINEGQNKPIEVWAKQGBTERLCSQNPVVLKHQBREITRTTLQSDQEB 1680
DB 1621 NACESNHAIAINEGQNKPIEVWAKQGBTERLCSQNPVVLKHQBREITRTTLQSDQEB 1680
QY 1681 IDYDDTIISVEMKKEDFDIYEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
DB 1681 IDYDDTIISVEMKKEDFDIYEDENQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
QY 1741 AQSGSVQPKVQVFEFTDGSFTQPLYRGELNDELGLLGPYIRAEVEDNIWVTFRNQASR 1800
DB 1741 AQSGSVQPKVQVFEFTDGSFTQPLYRGELNDELGLLGPYIRAEVEDNIWVTFRNQASR 1800
QY 1801 PYSFYSSLISYEDQRCQAEPRKFNFKVQNHMAPTKDFDCKAWAFSDV 1860

DB 1801 PYSFYSSLISYEDQRCQAEPRKFNFKVQNHMAPTKDFDCKAWAFSDV 1860
QY 1861 DLEKDVHGLIGPLLVCHTNTINPAHGQVTVQEZALPFTTFDETKSWYFTFENMERNCA 1920
DB 1861 DLEKDVHGLIGPLLVCHTNTINPAHGQVTVQEZALPFTTFDETKSWYFTFENMERNCA 1920
QY 1921 PCNIQMEDEPTKENVRFHAINGYIMDTLPGLVMAOQRIWYLLSMGNGENIHSIHFSGH 1980
DB 1921 PCNIQMEDEPTKENVRFHAINGYIMDTLPGLVMAOQRIWYLLSMGNGENIHSIHFSGH 1980
QY 1981 VFTVRKXBEYKXNALXLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGNSTLFLVYSNKC 2040
DB 1981 VFTVRKXBEYKXNALXLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGNSTLFLVYSNKC 2040
QY 2041 QTPLEGVASHIRDFQITASGOYQOKAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMII 2100
DB 2041 QTPLEGVASHIRDFQITASGOYQOKAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMII 2100
QY 2101 HGIKTQGARQKFESSLYISQFIIMVSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIFN 2160
DB 2101 HGIKTQGARQKFESSLYISQFIIMVSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIFN 2160
QY 2161 PPIIARYIRLHPHTHYSIRSTLRMELMGCGLNCSMPLGWESKAI SDAQITASSYFTNMPA 2220
DB 2161 PPIIARYIRLHPHTHYSIRSTLRMELMGCGLNCSMPLGWESKAI SDAQITASSYFTNMPA 2220
QY 2221 TWSPSKARLHLQGRNANWEPQVNNPEKMLQVDFOKTMKVGTTCQGVKSLTSMYVKFSL 2280
DB 2221 TWSPSKARLHLQGRNANWEPQVNNPEKMLQVDFOKTMKVGTTCQGVKSLTSMYVKFSL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDQSFPTPVVNSLDPLLLTRYLRIHPQSWVHQIALRM 2340
DB 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDQSFPTPVVNSLDPLLLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EYLGEBAQDLY 2351
DB 2341 EYLGEBAQDLY 2351
RESULT 3
AAW46245
ID AAW46245 standard; protein; 2351 AA.
XX AAW46245;
AC XX
DT 25-MAR-2003 (revised)
DT 06-AUG-1998 (first entry)
XX XX
DE Human factor III protein sequence.
KW Replication defective; recombinant retrovirus; RRV; therapeutic protein;
KW haemophilia; thrombosis; hypercoagulable disorder; liver disease; human;
KW hepatitis; thalassemia; phenylketonuria; Leisch-Nyhan syndrome; diabetes;
KW cystic fibrosis; Duchenne's Muscular Dystrophy; hypercholesterolemia;
KW hypopituitarism; adenine deaminase deficiency; HIV infection; anaemia;
KW Guacher's syndrome; high blood pressure; Alzheimer's disease, factor III;
KW autoimmune; inflammatory disease.
XX XX
OS Homo sapiens.
XX XX
FN WO9800541-A2.
PD 08-JAN-1998.
PF 02-JUL-1997; 97WO-US011784.
XX 03-JUL-1996; 96US-00645601.
PR 13-AUG-1996; 96US-00696381.
PR 04-JUN-1997; 97US-00869309.
XX XX
PA (CHIR) CHIRON CORP.
XX XX

PI Jolly DU, Barber JR, Chang SMW, Respass JG, Allen JR, Boder M;
 PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanes CE;
 PI Mittelstaedt DM, Prussak CE, Greengard J, Lee R;
 XX
 DR WPI: 1998-086966/08.
 DR N-PSDB; AAV19580.

XX New replication defective recombinant retro-viruses - which can be
 PT administered to provide long term systemic expression of therapeutic
 PT protein in blood, useful in, e.g. treating hyper-coagulable disorders.
 XX
 XX Example 28; Page 203-210; 272pp; English.

XX This is the human factor III sequence. The encoding DNA is used in the
 CC construction of recombinant retroviral vectors expressing human factor
 CC VIII. The specification provides the preparation of replication defective
 CC recombinant retrovirus (RRV) expressing a therapeutic protein. The RRV
 CC preparation is resistant to degradation by human complement and is
 CC capable of inducing long term systemic expression of the therapeutic
 CC protein when administered intravenously to a human. The long term
 CC systemic expression results in a measurable level of the therapeutic
 CC protein being produced in the blood of the human for a period of at least
 CC 30 days after the administration of the RRV vector preparation. RRV's can
 CC be used for in vivo delivery of therapeutic protein to treat, e.g.
 CC haemophilia A, haemophilia B, thrombosis, hypercoagulable disorders,
 CC liver diseases such as hepatitis, disorders such as thalassemia,
 CC phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency
 CC (SCID), cystic fibrosis, Duchenne's Muscular Dystrophy, inherited
 CC emphysema, familial hypercholesterolemia, diabetes, hypopituitarism,
 CC adenine deaminase deficiency, algal-antitrypsin deficiency, Guacher's
 CC syndrome, anaemia, infections such as HIV infection, high blood pressure,
 CC Alzheimer's disease, autoimmune or inflammatory disease or graft versus
 CC host disease. RRV's are capable of surviving inactivation in human serum
 CC thereby allowing efficient gene transfer over prolonged periods of time.
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 2351 AA;

Query Match 99.9%; Score 12399; DB 2; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQIELSTCFFLCFLRCPFSATRRYYLGAVELSWYQSDGLGELPVDARPPRPVPSKFFPN 60
 DB 1 MQIELSTCFFLCFLRCPFSATRRYYLGAVELSWYQSDGLGELPVDARPPRPVPSKFFPN 60
 QY 61 TSVVYKTLFVEFTDHLNIAKPRPPWMLGGLPTTQAEVDTWVITLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLNIAKPRPPWMLGGLPTTQAEVDTWVITLKNMASHPVSLHAV 120
 QY 121 GVSYWKASGAEYDDQTSOREKEDDKVPFGGSHTYVWVLKENGWASDPLCLTYSYLSH 180
 DB 121 GVSYWKASGAEYDDQTSOREKEDDKVPFGGSHTYVWVLKENGWASDPLCLTYSYLSH 180
 QY 181 VDLVKDINSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
 DB 181 VDLVKDINSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
 QY 241 AASARAWPKHTVNGVYVNRSLPGLIGCHRKSVYWHVIGMTTPVHVSIFLEGHTPLVRNH 300
 DB 241 AASARAWPKHTVNGVYVNRSLPGLIGCHRKSVYWHVIGMTTPVHVSIFLEGHTPLVRNH 300
 QY 301 RQASLEISPIITFLTAQTLMDLQFLLSCHISSHQHDMGEAVVKYVDSQPEEPQIMKXNE 360
 DB 301 RQASLEISPIITFLTAQTLMDLQFLLSCHISSHQHDMGEAVVKYVDSQPEEPQIMKXNE 360
 QY 361 EADYDDDLTDEMDVVRPDDNNSPSFQIRSVAKKHPTVHYIAAEEEDWDYAPLVIA 420
 DB 361 EADYDDDLTDEMDVVRPDDNNSPSFQIRSVAKKHPTVHYIAAEEEDWDYAPLVIA 420
 QY 421 PDRSVKQYLNNGPQIRGRKVKYRMAVDTETFKTRAQCHESGILGPLLYGVBGDTL 480
 DB 421 PDRSVKQYLNNGPQIRGRKVKYRMAVDTETFKTRAQCHESGILGPLLYGVBGDTL 480

QY 481 LIIFKNQASRPNIYFPHGIDTVRPLYSRRLPKGVKHLKDFPILPGEIRPKYKWTVTVEDGP 540
 DB 481 LIIFKNQASRPNIYFPHGIDTVRPLYSRRLPKGVKHLKDFPILPGEIRPKYKWTVTVEDGP 540
 QY 541 TKSDPRCLTRYVSSFVNWERDLASGLIGPLLCYKESVDKGNQIMSKRNVILFSVDE 600
 DB 541 TKSDPRCLTRYVSSFVNWERDLASGLIGPLLCYKESVDKGNQIMSKRNVILFSVDE 600
 QY 601 NRSWYLITENIQRFLEPNPAGVQLEDPEFQASNMHSINGVYVDFSLQSLVCLHEVAYWYLS 660
 DB 601 NRSWYLITENIQRFLEPNPAGVQLEDPEFQASNMHSINGVYVDFSLQSLVCLHEVAYWYLS 660
 QY 661 IGAQTDPLSVFSGVTFKHKWYEDTILFPFSGHTVFMENPGLWILGCHNSFRNRG 720
 DB 661 IGAQTDPLSVFSGVTFKHKWYEDTILFPFSGHTVFMENPGLWILGCHNSFRNRG 720
 QY 721 MTALLKVSSCDKNTGDYVEDSVEDISAYLLSKNNAIEPRSPNSRRHPSRTRQKQFNATTI 780
 DB 721 MTALLKVSSCDKNTGDYVEDSVEDISAYLLSKNNAIEPRSPNSRRHPSRTRQKQFNATTI 780
 QY 781 PENDIEKTDMPFAHRTMPKIQNVSSDMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840
 DB 781 PENDIEKTDMPFAHRTMPKIQNVSSDMLLRQSPTPHGLSLSDLOEAKYETFSDDPS 840
 QY 841 PGADSNNSLSEMTFRPOLHSCDMVTPESGLQRLNEKLGTTAAATELKLPKVSST 900
 DB 841 PGADSNNSLSEMTFRPOLHSCDMVTPESGLQRLNEKLGTTAAATELKLPKVSST 900
 QY 901 SNNLITIPSDNLAAGTNTSSLPSPMPVHYDSQDPTTLFGKSSSPITESGGPLSLEE 960
 DB 901 SNNLITIPSDNLAAGTNTSSLPSPMPVHYDSQDPTTLFGKSSSPITESGGPLSLEE 960
 QY 961 NNDKLLBSGLMNGQESWGKNVSTESGRLPFKGRAHGPALLTKDNALFKVSI SLAKTN 1020
 DB 961 NNDKLLBSGLMNGQESWGKNVSTESGRLPFKGRAHGPALLTKDNALFKVSI SLAKTN 1020
 QY 1021 KTSNNSATNRKTHIDGSLLLJENSVPWONLLESTTEKKVTPLIHDRLMDKNAALRL 1080
 DB 1021 KTSNNSATNRKTHIDGSLLLJENSVPWONLLESTTEKKVTPLIHDRLMDKNAALRL 1080
 QY 1081 NMSNKTTSKKNMVMVQKKEGPVPPAQNPDMFPPKMLFLPESARWQRTGHKNSLNSG 1140
 DB 1081 NMSNKTTSKKNMVMVQKKEGPVPPAQNPDMFPPKMLFLPESARWQRTGHKNSLNSG 1140
 QY 1141 QGPSFKQLVSLGPKSVGQNFSEKKNVVGKGEFTKDVGLKEMVFPSSRNLFUTNLDN 1200
 DB 1141 QGPSFKQLVSLGPKSVGQNFSEKKNVVGKGEFTKDVGLKEMVFPSSRNLFUTNLDN 1200
 QY 1201 LHENNTNQEKKIOEEIEKKETLIQENVVLPQIHTVTGTKNPMKNLFLLSRQNVGSD 1260
 DB 1201 LHENNTNQEKKIOEEIEKKETLIQENVVLPQIHTVTGTKNPMKNLFLLSRQNVGSD 1260
 QY 1261 GAYAPVLQDPRSLANDSTNRKTKHTAHFSPKSGEENLEGLNQTKQIVEKYACTRISENT 1320
 DB 1261 GAYAPVLQDPRSLANDSTNRKTKHTAHFSPKSGEENLEGLNQTKQIVEKYACTRISENT 1320
 QY 1321 SQOQFVTVORSKRALQKQPLPEETLEKRIITVDDTSTQWSKNMKHLTFTSTLTQDYNEKE 1380
 DB 1321 SQOQFVTVORSKRALQKQPLPEETLEKRIITVDDTSTQWSKNMKHLTFTSTLTQDYNEKE 1380
 QY 1381 KGATITQSPLSCLTRSHSIPOANRSPPIAKVSSFPSPIRPIYLRVLFDQNSHLPAAASY 1440
 DB 1381 KGATITQSPLSCLTRSHSIPOANRSPPIAKVSSFPSPIRPIYLRVLFDQNSHLPAAASY 1440
 QY 1441 RKXDSGVQESSHFLQGAKKNNLSLAITLMTGQREYVSLGTSATNSVTYKVKVENTVLP 1500
 DB 1441 RKXDSGVQESSHFLQGAKKNNLSLAITLMTGQREYVSLGTSATNSVTYKVKVENTVLP 1500
 QY 1501 KPDLPKTSKGKVELLPKHVYIYOKDLFPPTETSNVSGSPGHLDLVEGSLQGTGEGAIKWNEANRP 1560
 DB 1501 KPDLPKTSKGKVELLPKHVYIYOKDLFPPTETSNVSGSPGHLDLVEGSLQGTGEGAIKWNEANRP 1560


```
QY 1561 GKVPFLRVATESSAKTPSKLLDPLANDNHVGTQIPKEEWSQKSPKTAFAKKKDTILSL 1620
Db 1561 GKVPFLRVATESSAKTPSKLLDPLANDNHVGTQIPKEEWSQKSPKTAFAKKKDTILSL 1620
QY 1621 NACESNHAIANAINEGQNKPIEYTWAKQRTHELKCSQNPVVKHQREIRTTQLSDQEE 1680
Db 1621 NACESNHAIANAINEGQNKPIEYTWAKQRTHELKCSQNPVVKHQREIRTTQLSDQEE 1680
QY 1681 IDYDDTISVEMKEDDFDIYEDENQSPRSFQKTRHYFIAAVERLDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKEDDFDIYEDENQSPRSFQKTRHYFIAAVERLDYGMSSSPHVLNR 1740
QY 1741 AQSGVPFQKVVQFQFTDGSFTQPLVRGELNHELGLLGFYIRAEVEDNIMVTRNQASR 1800
Db 1741 AQSGVPFQKVVQFQFTDGSFTQPLVRGELNHELGLLGFYIRAEVEDNIMVTRNQASR 1800
QY 1801 PYSFYSLISYEDORQGAEPKRNFKYKPNKTKYFWKVQHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSLISYEDORQGAEPKRNFKYKPNKTKYFWKVQHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGLIGLPLVCHTNTLPAHGRQVTVQEFALFETIPDETSSWYFTENVEENCR 1920
Db 1861 DLEKDVHSGLIGLPLVCHTNTLPAHGRQVTVQEFALFETIPDETSSWYFTENVEENCR 1920
QY 1921 PCNIQMEDPTFKENYRFAHNGYIMDTPLGLVVAQDQIRWYLLSMGNSNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFAHNGYIMDTPLGLVVAQDQIRWYLLSMGNSNENIHSIHFSGH 1980
QY 1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGHLHAGHSTLELVTSNKC 2040
Db 1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKAGIWRVECLIGHLHAGHSTLELVTSNKC 2040
QY 2041 QTEPLGASGHIRDFQITASQYQWAPKPLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTEPLGASGHIRDFQITASQYQWAPKPLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
QY 2101 HGKTCQAKQKFSLSIYISQPIIMYSIDGKKWQYRGNSTGTMVFFGNVDSSGIKENIFN 2160
Db 2101 HGKTCQAKQKFSLSIYISQPIIMYSIDGKKWQYRGNSTGTMVFFGNVDSSGIKENIFN 2160
QY 2161 PPIIARYIRLHPTHYSIRSLRMLGCDLNSCMLPGMESKAIQAQITASSYFTNMPA 2220
Db 2161 PPIIARYIRLHPTHYSIRSLRMLGCDLNSCMLPGMESKAIQAQITASSYFTNMPA 2220
QY 2221 TWSPSKARLHLOGRSNARPOVNNPKWLQVDQPKTKMKTGVTQGVKSLITSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNARPOVNNPKWLQVDQPKTKMKTGVTQGVKSLITSMYVKEFL 2280
QY 2281 ISSQDGHQWTLFPQNGKVKVFGNQDSFTPVVNSLDPPLTRYLRIHPQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFPQNGKVKVFGNQDSFTPVVNSLDPPLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351
```

RESULT 4

```
AAW44373
ID AAW44373 standard; protein; 2351 AA.
AC AAW44373;
XX
XX
XX 20-JUL-1998 (first entry)
DT Human Factor VIII.
DE Factor VIII; blood clotting; haemophilia A; gene therapy; retrovirus;
KW vector; human.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
```

```
Domain 711..1694
/notes="B-domain"
PN W09800542-A2.
XX
XX 08-JAN-1998.
XX
XX 02-JUL-1997; 97NO-US011785.
XX
XX 03-JUL-1996; 96US-00545601.
PR 13-AUG-1996; 96US-00596381.
PR 04-JUN-1997; 97US-00369309.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Jolly DJ, Barber JR, Chang SM, Respass JG, Allen JR, Boder M;
PI Chong K, De La Vega D, Depolo NJ, Hsu DC, Ibanez CE;
PI Mittelstaedt DM, Prussak CE, Greengard J;
XX
XX WPI: 1998-086967/08.
DR N-PSDB; AAV15359.
XX
XX New replication defective recombinant retroviruses - which express B
PI domain-deleted human factor VIII or human factor IX for the treatment of
XX haemophilia.
XX
XX Example 28; Page 164-166; 236pp; English.
XX
XX This polypeptide comprises human Factor VIII. The invention relates to
CC preparations of replication defective recombinant retrovirus (RV)
CC expressing a B domain-deleted human Factor VIII protein, such as the SQN
CC deletion mutant (see AAW44372), where the recombinant RV is capable of
CC infecting human cells, is resistant to degradation by human complement
CC and is capable of inducing long-term (at least 30 days and up to 6 months
CC or longer post-injection) systemic expression of Factor VIII when
CC administered to a haemophilia A patient
XX
XX Sequence 2351 AA;
Query Match 99.9%; Score 12399; DB 2; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOELSTCFCLLAFCTSATRRYVLGAVELSWDWCQSDGLBPLVDARPPVPKSPFN 60
Db 1 MOELSTCFCLLAFCTSATRRYVLGAVELSWDWCQSDGLBPLVDARPPVPKSPFN 60
QY 61 TSVVYKTLFVFTDELFNIAKPRPPWMLGLOPTTQAEVYDTVITLKNMASHPVSLHAV 120
Db 61 TSVVYKTLFVFTDELFNIAKPRPPWMLGLOPTTQAEVYDTVITLKNMASHPVSLHAV 120
QY 121 GVSVMKASGAEYDDQTSQREKDDKVPFGSHTYVQVLKENGMSADPLCLTYSYLSH 180
Db 121 GVSVMKASGAEYDDQTSQREKDDKVPFGSHTYVQVLKENGMSADPLCLTYSYLSH 180
QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLKFIILLFAVDEGKSWSEKTNLSMQDRD 240
Db 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLKFIILLFAVDEGKSWSEKTNLSMQDRD 240
QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELCAQTLLMDLQFLLSCHTSSHQHDCGMEAVKYVDSCEPSPQLMKNE 360
Db 301 ROASLEISPIITELCAQTLLMDLQFLLSCHTSSHQHDCGMEAVKYVDSCEPSPQLMKNE 360
QY 361 EAEDYDDDLTDSMDVVRPDDNDSPEFQIRSVAKKHPTWVHYIAAEEEDWDYAPLVIA 420
Db 361 EAEDYDDDLTDSMDVVRPDDNDSPEFQIRSVAKKHPTWVHYIAAEEEDWDYAPLVIA 420
QY 421 PDRSYKSOYLNGPQIRGRKYKVRPMAYTDETPKTEAIOHRSGLIGPLLYGBVGTLL 480
Db 421 PDRSYKSOYLNGPQIRGRKYKVRPMAYTDETPKTEAIOHRSGLIGPLLYGBVGTLL 480
```

Db 421 PDRSYKSYQLNNGPQRIGRKYKVRPMAYTDTEFTKTRAIQHESGILGPLLYGEVGDTL 480
QY 481 LIIFKNQASRPYNIYEHGIDVRLPRLSRRLPKGVKHLKDPILPGETFKYKWTVTVEDGP 540
Db 481 LIIFKNQASRPYNIYEHGIDVRLPRLSRRLPKGVKHLKDPILPGETFKYKWTVTVEDGP 540
QY 541 TKSDPRLCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQGNQIKXSDKENVILFSVFDE 600
Db 541 TKSDPRLCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQGNQIKXSDKENVILFSVFDE 600
QY 601 NRSWYLCTENIQRLEPNPAGVQLDPDPFOASNIHMSINGYVFDLSLOISVCLHEVAYWYLS 660
Db 601 NRSWYLCTENIQRLEPNPAGVQLDPDPFOASNIHMSINGYVFDLSLOISVCLHEVAYWYLS 660
QY 661 IGAQTDFLSVFFSGYTFKHKWYEDTLTLPFGGETVFMGMENPGLWILGCNSDPRNRG 720
Db 661 IGAQTDFLSVFFSGYTFKHKWYEDTLTLPFGGETVFMGMENPGLWILGCNSDPRNRG 720
QY 721 MTALLKVSSCDKNTGYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTQKQFNATTI 780
Db 721 MTALLKVSSCDKNTGYEDSYEDISAYLLSKNAIEPRSFQNSRHPSTQKQFNATTI 780
QY 781 PENDIEKTDWFAHRTMPMKIQNVSSDILLMLRQSTPHCLSLSDI:OBAKXITFSDDPS 840
Db 781 PENDIEKTDWFAHRTMPMKIQNVSSDILLMLRQSTPHCLSLSDI:OBAKXITFSDDPS 840
QY 841 PGADSNNSLSEMTFRPOLHSGDMVFTPEGLQLRLNEKLTAAATELKKLDFKVSST 900
Db 841 PGADSNNSLSEMTFRPOLHSGDMVFTPEGLQLRLNEKLTAAATELKKLDFKVSST 900
QY 901 SNNLSTIFSDNLAAGTDNTSSIGPSPMPVHYDSQDITLFGKSSPFTESGGPLSLSEE 960
Db 901 SNNLSTIFSDNLAAGTDNTSSIGPSPMPVHYDSQDITLFGKSSPFTESGGPLSLSEE 960
QY 961 NNDSKLLLEGSLMNSQSSGKNVSSSTESGRLPKGAHGPALLTKONALFKVISILKTN 1020
Db 961 NNDSKLLLEGSLMNSQSSGKNVSSSTESGRLPKGAHGPALLTKONALFKVISILKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDETFEKVTPLIHDRMLMDKNATALRL 1080
Db 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDETFEKVTPLIHDRMLMDKNATALRL 1080
QY 1081 NKSNNKTTSSKNMVMQOKKEGIPDPAQNDMSFFKMLFPESAAMIORTHGKNSLNSG 1140
Db 1081 NKSNNKTTSSKNMVMQOKKEGIPDPAQNDMSFFKMLFPESAAMIORTHGKNSLNSG 1140
QY 1141 QGSPKQVLSLGPKEVEGONFLSEKNKVVVGKGEFTKQGLKEMVFPSSRNLFLTNDN 1200
Db 1141 QGSPKQVLSLGPKEVEGONFLSEKNKVVVGKGEFTKQGLKEMVFPSSRNLFLTNDN 1200
QY 1201 LHENNTNHOEKKI:QBEIERKKTILIQENVVLPOIHTVTGTFKNFMKNLFLASTRQNVGSYD 1260
Db 1201 LHENNTNHOEKKI:QBEIERKKTILIQENVVLPOIHTVTGTFKNFMKNLFLASTRQNVGSYD 1260
QY 1261 GAYAPVLQDPRSLNSTNTEKKTAAHFSKKGREENLEGLGNQTKQIVEKYACTTRISPT 1320
Db 1261 GAYAPVLQDPRSLNSTNTEKKTAAHFSKKGREENLEGLGNQTKQIVEKYACTTRISPT 1320
QY 1321 SQONFVTRASKRALKOFLPLETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDVNEKE 1380
Db 1321 SQONFVTRASKRALKOFLPLETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDVNEKE 1380
QY 1381 KGAIQTSPSLDCLTRSHSIPQANRPLPIAKVSSPSPRIPIYLTRVLFQDNSSHLPAAASY 1440
Db 1381 KGAIQTSPSLDCLTRSHSIPQANRPLPIAKVSSPSPRIPIYLTRVLFQDNSSHLPAAASY 1440
QY 1441 RKKDSGVQSSSHFLOGAKKNLSLAILTLEMTGDQREVGLSATSNTSVYTKKVENTVLP 1500
Db 1441 RKKDSGVQSSSHFLOGAKKNLSLAILTLEMTGDQREVGLSATSNTSVYTKKVENTVLP 1500
QY 1501 KPDLFKTSKVELLPKWHYQKDLFPETESNGSPGHLDLVEGSL:CGTFEGALKWNEANRP 1560
Db 1501 KPDLFKTSKVELLPKWHYQKDLFPETESNGSPGHLDLVEGSL:CGTFEGALKWNEANRP 1560

QY 1561 CKVPPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQBSPEKTAFKKDKTILSL 1620
Db 1561 CKVPPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEEWSQBSPEKTAFKKDKTILSL 1620
QY 1621 NACESNHAAANEGONKPEIEVWAKQGRTERLCSNPPVLKHHOREITRTTTLQSDQEE 1680
Db 1621 NACESNHAAANEGONKPEIEVWAKQGRTERLCSNPPVLKHHOREITRTTTLQSDQEE 1680
QY 1681 IDYDDTISVEMKCKEDFIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKCKEDFIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AOSGSVPOKKVYFQBPSTGSTOPLYRGELNEHLGLGPYTRAEVEDIMVTFRNOASR 1800
Db 1741 AOSGSVPOKKVYFQBPSTGSTOPLYRGELNEHLGLGPYTRAEVEDIMVTFRNOASR 1800
QY 1801 PYSFYSSLLSYBEDQOQAEPRKGFVKPNETKTYPKVQVHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSSLLSYBEDQOQAEPRKGFVKPNETKTYPKVQVHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKOVHSGLIGPLAVCHTNTLNPAHGRQVTVQEPALFTIETDKSWYFTENMERNCRA 1920
Db 1861 DLEKOVHSGLIGPLAVCHTNTLNPAHGRQVTVQEPALFTIETDKSWYFTENMERNCRA 1920
QY 1921 PCNIQMEDPTKENYRFFHAINGYIMDTPLGLVMAQDQRIRWYLLSMGSHENIHS:HFSGH 1980
Db 1921 PCNIQMEDPTKENYRFFHAINGYIMDTPLGLVMAQDQRIRWYLLSMGSHENIHS:HFSGH 1980
QY 1981 VFTVRKKEYSKMALNLYPGVFETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEYSKMALNLYPGVFETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKC 2040
QY 2041 QTPFLGWSAGHIREDPQITASGGYQGNAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPFLGWSAGHIREDPQITASGGYQGNAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
QY 2101 HGIKTOGARQKFSYLSISQFIIMVSLDGKKWOTYKGNSTGTLMPFGNVDSSGIKHNIN 2160
Db 2101 HGIKTOGARQKFSYLSISQFIIMVSLDGKKWOTYKGNSTGTLMPFGNVDSSGIKHNIN 2160
QY 2161 PPIIARYIRLHPTHYSIRSTLEMLMGCDLNSCSMPLGMESKAI:SDAQITASSYFTNMEA 2220
Db 2161 PPIIARYIRLHPTHYSIRSTLEMLMGCDLNSCSMPLGMESKAI:SDAQITASSYFTNMEA 2220
QY 2221 TWSPKARLHLOGRSNARPPQVNNPKMLQVDFQKTMKVGTGTTQGVKSLLTSMVYKEPL 2280
Db 2221 TWSPKARLHLOGRSNARPPQVNNPKMLQVDFQKTMKVGTGTTQGVKSLLTSMVYKEPL 2280
QY 2281 ISSSODGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPLLTRYLRIHPQSWVHQAIALRM 2340
Db 2281 ISSSODGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPLLTRYLRIHPQSWVHQAIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351
RESULT 5
ID AAY52537 standard; protein; 2351 AA.
XX AAY52537;
AC AAY52537;
XX
DT 28-FEB-2000 (first entry)
XX Human full-length factor VIII.
DE Factor VIII; haemophilia; proteolysis; heavy chain; light chain;
KW secretion; von Willebrand Factor; vWF; C2 domain; intron 22; inversion;
XX non-functional; bleeding disorder; coagulation; treatment.
XX Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..19
 FT /note= "Signal peptide"
 FT Protein 20..1708
 FT /note= "Maximum length human factor VIII heavy chain"
 FT Protein 1709..2351
 FT /note= "Human factor VIII light chain"
 XX WO959622-A1.
 XX 25-NOV-1999.
 XX 17-MAY-1999; 99WO-US010872.
 XX 20-MAY-1999; 98US-00092000.
 XX (GETH) GENENTECH INC.
 FA Vohar GA;
 XX WPI; 2000-053195/04.
 DR N-PSDB; AA238604.
 XX New method for treating hemophilia A.
 PT Disclosure; Fig 9; 50pp; English.
 XX
 XX This sequence represents the full-length human factor VIII. In this form, the protein is 300 kD in size with the domain structure A1-A2-B-A3-C1-C2. However, prior to secretion, this 300 kD protein is proteolyzed into a heavy chain (A1-A2-B, with continued proteolysis within the B domain resulting in molecules of varying length) and a light chain (A3-C1-C2) that remains non-covalently attached to the heavy chain. Upon secretion, factor VIII is rapidly cleared from the circulation unless it is bound by the plasma protein von Willebrand Factor (vWF). Factor VIII binds to vWF through the light chain, with known binding sites at the N-terminus and within the C2 domain. Haemophilia A is frequently caused by an intron 22 inversion in the factor VIII gene, which separates the gene into two sections, exons 1-22 becoming inverted and localised telomeric to the original site, while exons 23-26 remain isolated at the original site. Exons 23-26 a portion of the C1 and all of the C2 domains, without which factor VIII is non-functional. The factor VIII gene product of individuals with this mutation thus comprises domains A1-A2-B-A3 plus a fragment of C1, which on proteolysis is non-functional, resulting in a bleeding disorder. The invention relates to a novel method for for treating haemophilia A in a mammal able to produce the factor VIII heavy chain. The method comprises administering to the mammal either an effective amount of factor VIII light chain, or a gene encoding it, and may be useful for treating patients such as those whose haemophilia A is caused by intron 22 inversion. The recombinant factor VIII products of this invention are derived from well-characterised starting materials which therefore reduces the risk of pathogenic infection which was previously a problem when using donated plasma. Furthermore, the invention provides a more economic and potentially more effective treatment for haemophilia. There is also a need of providing factor VIII activity to patients who produce or are at risk of producing antibodies against full-length factor VIII
 XX Sequence 2351 AA;
 SQ
 Query Match 99.9%; Score 12399; DB 3; Length 2351;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MQELSTCFPLCLLRCEFSATREYVYLGAVELSDWYMQSDLGELPVDARPPVVKSPFEN 60
 DB 1 MQELSTCFPLCLLRCEFSATREYVYLGAVELSDWYMQSDLGELPVDARPPVVKSPFEN 60
 QY 61 TSVYVKTLFVEFTDHLFNIAKPRPPWMLGPGTIOAEVYDVTVITLKNWASHFVSLHAV 120
 DB 61 TSVYVKTLFVEFTDHLFNIAKPRPPWMLGPGTIOAEVYDVTVITLKNWASHFVSLHAV 120

QY 121 GVS YWKASGAEYDDOTSORKEKDDKVPGGSHYVVOVLKENGWASDPLCLCTSYLSH 180
 DB 121 GVS YWKASGAEYDDOTSORKEKDDKVPGGSHYVVOVLKENGWASDPLCLCTSYLSH 180
 QY 181 VDLVKDNLGSLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQORD 240
 DB 181 VDLVKDNLGSLIGALLVCREGSLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQORD 240
 QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
 DB 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
 QY 301 ROASLEISPIETELTAQTLLMDLQPLLSCHEISSHODGMEAVKVVDSPERPEKMKNE 360
 DB 301 ROASLEISPIETELTAQTLLMDLQPLLSCHEISSHODGMEAVKVVDSPERPEKMKNE 360
 QY 361 EADYDDDLTDSMDVWRFDDNPSFIIQIRSVAKKHFKTWVHYIAAEEEDWDYAPLVLA 420
 DB 361 EADYDDDLTDSMDVWRFDDNPSFIIQIRSVAKKHFKTWVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDRSYKSOYLNNGPQRIGRKYKVRPMAYTDETFKTRAIQHESGILGPLYGEVDTL 480
 DB 421 PDRSYKSOYLNNGPQRIGRKYKVRPMAYTDETFKTRAIQHESGILGPLYGEVDTL 480
 QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVVEQCP 540
 DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIPKYKWTVVEQCP 540
 QY 541 TKSDPRCLTRYSSPVNMRDLASGLIGPLLCYKESVDKGNQIMSKRNVLFSVDE 600
 DB 541 TKSDPRCLTRYSSPVNMRDLASGLIGPLLCYKESVDKGNQIMSKRNVLFSVDE 600
 QY 601 NESWYLTENIOFLNPAGVOLDEPFOASNIWHGINSVVDLSIOLSVCLHEVAWYILS 660
 DB 601 NESWYLTENIOFLNPAGVOLDEPFOASNIWHGINSVVDLSIOLSVCLHEVAWYILS 660
 QY 661 IGAQDIFLSVFPFGYTFKHKWVYEDTLTFPPSGETVPMSENPGLWILGCHNSFRNRG 720
 DB 661 IGAQDIFLSVFPFGYTFKHKWVYEDTLTFPPSGETVPMSENPGLWILGCHNSFRNRG 720
 QY 721 MTALLKVSCKDNGDYEDSVEDISAVLLSKNNAIEPRFSQNSRHPSTROKQNATTI 780
 DB 721 MTALLKVSCKDNGDYEDSVEDISAVLLSKNNAIEPRFSQNSRHPSTROKQNATTI 780
 QY 781 PENDIEKTDPEFAHTEPMPKIONVSSDLMLLRQSPHGLSLDLOBAKYETESDDPS 840
 DB 781 PENDIEKTDPEFAHTEPMPKIONVSSDLMLLRQSPHGLSLDLOBAKYETESDDPS 840
 QY 841 PGAIIDNNLSSEMTFRPQLHSGDMVFTPESSGLQRLNEKLGTTAATLKKLDFKVSST 900
 DB 841 PGAIIDNNLSSEMTFRPQLHSGDMVFTPESSGLQRLNEKLGTTAATLKKLDFKVSST 900
 QY 901 SNNLSTIPSDMLAAGTNTSSLGPPMPVHYVDSOLDTTLFGKKSPLTESGGPLSLBEE 960
 DB 901 SNNLSTIPSDMLAAGTNTSSLGPPMPVHYVDSOLDTTLFGKKSPLTESGGPLSLBEE 960
 QY 961 NNDKLLBSGLMNSQESSGKNVSVSTESGRLPKGRAGHPALLTKDNALFKVSIILLKTN 1020
 DB 961 NNDKLLBSGLMNSQESSGKNVSVSTESGRLPKGRAGHPALLTKDNALFKVSIILLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSPPVMQNILESDETFKKVTPLIHDRMLMDKNATALRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPPVMQNILESDETFKKVTPLIHDRMLMDKNATALRL 1080
 QY 1081 NEMSNKTTSSKNMEXVQKKEGPIPPDAQNPDMSFFKMLFLPESARWQRTHQKNSLNSG 1140
 DB 1081 NEMSNKTTSSKNMEXVQKKEGPIPPDAQNPDMSFFKMLFLPESARWQRTHQKNSLNSG 1140
 QY 1141 QGPSPKQLVSLGPEKSVGQNPFLSEKNKVVGKGFETKDVGLKEMVFPSSRNLFITNLDN 1200
 DB 1141 QGPSPKQLVSLGPEKSVGQNPFLSEKNKVVGKGFETKDVGLKEMVFPSSRNLFITNLDN 1200
 QY 1201 LHENNTNQEKKIQEEIEKKETLIQENVVLPOIHTVTGTQNFNMKNLFLSTRQNVGSYD 1260

Db 1201 LHENNNHQQEKKIQSEIEBKKEKTLIQENVVLPQIHVTGKPMKNI.FLLSTRQNVGSDYD 1260
QY 1261 GAYAVLQDFSLNDSNTRTKKHTAHFSPKGBEENLBGLGNQTKQIVKVIACITRISNT 1320
Db 1261 GAYAVLQDFSLNDSNTRTKKHTAHFSPKGBEENLBGLGNQTKQIVKVIACITRISNT 1320
QY 1321 SQONFVQSRKALQFRLPLETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQONFVQSRKALQFRLPLETELEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLPIAKVSSPFSIRPIYLTVLFDQDNSSHLPAASY 1440
Db 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLPIAKVSSPFSIRPIYLTVLFDQDNSSHLPAASY 1440
QY 1441 RKDQSGVQESSHFLQAGAKNNLSAILLTEMGDQREVGLSATSNTVYKKVENTVLP 1500
Db 1441 RKDQSGVQESSHFLQAGAKNNLSAILLTEMGDQREVGLSATSNTVYKKVENTVLP 1500
QY 1501 KPDLKPTSGKVELLPKVHIYQKDLPTFTSTNSPGHLDLVEGSLQGTGEGAIKWNENRP 1560
Db 1501 KPDLKPTSGKVELLPKVHIYQKDLPTFTSTNSPGHLDLVEGSLQGTGEGAIKWNENRP 1560
QY 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWSQEKSPKTAFFKKDITLSL 1620
Db 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEEWSQEKSPKTAFFKKDITLSL 1620
QY 1621 NACSENHAIANEGCNKPEIEVTWAKQRTRELCSQNPVPLKRCRREITRTTLOSQDEE 1680
Db 1621 NACSENHAIANEGCNKPEIEVTWAKQRTRELCSQNPVPLKRCRREITRTTLOSQDEE 1680
QY 1681 IDYDDTISVEMKKEFDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVEMKKEFDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSGVQVQPKVVFQFTDGSQTQPLRYGELNEHLGLGPIYRAEVENIMWTFRNQASR 1800
Db 1741 AQSGVQVQPKVVFQFTDGSQTQPLRYGELNEHLGLGPIYRAEVENIMWTFRNQASR 1800
QY 1801 PYSFYSSILSYEDORQGAEPKPNFVKNETTYFKVQOHMAPTKDFDCKAWAYFSDV 1860
Db 1801 PYSFYSSILSYEDORQGAEPKPNFVKNETTYFKVQOHMAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGLPLLVCHTNTLNPAHGRQVTVQEFALPFTIFDETKSWYFTENMERNCKA 1920
Db 1861 DLEKDVHSLGLPLLVCHTNTLNPAHGRQVTVQEFALPFTIFDETKSWYFTENMERNCKA 1920
QY 1921 PCNIQWEDPTPKENYRFHAINGIYMDTLPGLYMAQDQRIKRWYLLSGNSNENHSIHFSCH 1980
Db 1921 PCNIQWEDPTPKENYRFHAINGIYMDTLPGLYMAQDQRIKRWYLLSGNSNENHSIHFSCH 1980
QY 1981 VFTVRKKEBKALYNLYPGVFETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEBKALYNLYPGVFETVEMLPKAGIWRVECLIGEHLAGMSTLFLVYSNKC 2040
QY 2041 QTPLGWASGHIRDFQITAGSQYQWAPKLARLHYSGSINAWSTKPFWSIKVDLAPMII 2100
Db 2041 QTPLGWASGHIRDFQITAGSQYQWAPKLARLHYSGSINAWSTKPFWSIKVDLAPMII 2100
QY 2101 HGIKTQGAQKPSLLYSIOFIIMYSLDGKKWOTYRGNSTGLMWFGNVDSGSIKHNIEN 2160
Db 2101 HGIKTQGAQKPSLLYSIOFIIMYSLDGKKWOTYRGNSTGLMWFGNVDSGSIKHNIEN 2160
QY 2161 PPIIARYIRLHPHTHSIRSTRLMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPHTHSIRSTRLMELMGCDLNSCMPLGMESKAISDAQITASSYFTNMFA 2220
QY 2221 TWSPPSKARLHLQGRSNAPQVANNPKWLQVDPOKTMKVGTQGVTKSLLTSMYVKEFL 2280
Db 2221 TWSPPSKARLHLQGRSNAPQVANNPKWLQVDPOKTMKVGTQGVTKSLLTSMYVKEFL 2280
QY 2281 ISSODGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340

Db 2281 ISSODGHQWTLFFQNGKVKVQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAODLY 2351
Db 2341 EVLGCEAODLY 2351
RESULT 6
ABB99528
ID ABB99528 standard; protein; 2351 AA.
XX
AC ABB99528;
XX
DT 03-MAR-2003 (first entry)
XX
DE Amino acid sequence of human Factor VIII.
XX
KW Factor VIII; non-viral vesicle vector; vesicular membrane; hepatitis B; envelope protein; Factor IX; liver cell; gene therapy; haemophilia.
XX
OS Homo sapiens.
FN WO200286091-A2.
XX
PD 31-OCT-2002.
XX
PF 25-APR-2002; 2002WO-US013164.
XX
PR 25-APR-2001; 2001US-0286314P.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Chien KR, Hoshijima M;
DR WPI; 2003-093125/08.
DR N-PSDB; ABV77052.
XX
PT New nor-viral vesicle vector comprises vesicular membrane with hepatitis B envelope protein and nucleic acid expression construct comprising complete factor VIII or IX coding sequence, useful for treating hemophilia.
XX
PS Disclosure; Page 22-32; 34pp; English.
XX
The present sequence represents a human Factor VIII. Factor VIII polynucleotides are used to construct non-viral vesicle vectors. These vectors comprise a vesicular membrane with hepatitis B envelope protein exposed on the vesicle surface, and a nucleic acid expression construct comprising a complete factor VIII or factor IX coding sequence and a promoter sequence functional in liver cells. The non-viral vesicle vector is useful in gene therapy protocols for treating haemophilia
XX
SQ Sequence 2351 AA;
Query Match 99.9%; Score 12399; DB 5; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MOELSTCFFCLLPFCFSATRRYYLGAVELSWDVMQSDLGELPVDARPPRPVKSFPFN 60
Db 1 MOELSTCFFCLLPFCFSATRRYYLGAVELSWDVMQSDLGELPVDARPPRPVKSFPFN 60
QY 61 TSVYVKTLFVEFTDHLFNIAKPRPPMGLGPTTQAEVDTVTILKNMASHPVSLHAV 120
Db 61 TSVYVKTLFVEFTDHLFNIAKPRPPMGLGPTTQAEVDTVTILKNMASHPVSLHAV 120
QY 121 GVSYYWASGEGAEYDDQTSQRKEDDKVFPGSGHTYVWVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSYYWASGEGAEYDDQTSQRKEDDKVFPGSGHTYVWVLKENGPMASDPLCLTYSYLH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVDEGKSWHSETKNLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVDEGKSWHSETKNLMQDRD 240

RESULT 7

AAW11416
 ID AAW11416 standard; protein; 2351 AA.
 XX AAW11416;
 AC AAW11416;
 DT 20-NOV-1997 (first entry)
 XX
 DE Active Factor VIII:C analogue A774X.
 XX
 KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
 KW proteolytic cleavage.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH .19
 FT Peptide
 FT /note= "signal peptide"
 FT Protein
 FT 20..2351
 FT /note= "mature Factor VIII:C"
 FT Region
 FT 20..1667
 FT /note= "heavy chain fragment"
 FT Domain
 FT 760..1667
 FT /note= "B domain"
 FT Modified-site
 FT 774
 FT /label= phe, Glu, Pro
 FT Region
 FT 1668..2350
 FT /note= "light chain fragment"
 FT
 XX WO9703195-A1.
 XX
 PN 30-JAN-1997.
 PD
 PD
 XX 09-JUL-1996; 96WO-US011444.
 XX
 XX 11-JUL-1995; 95US-0001025P.
 PR
 XX (CHIR) CHIRON CORP.
 RA
 XX Hung DT, Cohen FE, Innis M;
 PI WPI; 1997-119050/11.
 XX
 DR Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 PT used in the treatment of haemophilias, by improvement of haemostasis.
 PT
 XX Claim 26; Page; 90pp; English.
 PS
 XX AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type factor
 CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 XX Sequence 2351 AA;

Query Match	99.8%	Score 12395;	DB 2;	Length 2351;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 2347;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;
QY 1	MOIELSTCFELCLLRFCSATRRYYIGAVELSDWYMQSDLGELPVDARFPFPRPKSFFPN	60		
DB 1	MOIELSTCFELCLLRFCSATRRYYIGAVELSDWYMQSDLGELPVDARFPFPRPKSFFPN	60		
QY 61	TSVYVKTLFVEFTDHLFNIAPRPWMLGLOPTTQAEVYDTVTITLKNMASHFVSLHAY	120		
DB 61	TSVYVKTLFVEFTDHLFNIAPRPWMLGLOPTTQAEVYDTVTITLKNMASHFVSLHAY	120		
QY 121	GVSYWKASGEAGYDDOTSQREKDDKVPFGSHTVWQVLKENGPMASDPLCLTYSYLH	180		
DB 121	GVSYWKASGEAGYDDOTSQREKDDKVPFGSHTVWQVLKENGPMASDPLCLTYSYLH	180		
QY 181	VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240		
DB 181	VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRD	240		
QY 241	AASARAWPMGHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEHTFLVRNH	300		
DB 241	AASARAWPMGHTVNGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEHTFLVRNH	300		
QY 301	ROASLEISPIITFLTAQTLLMDLGOFLLSCHISSHOHDGMEAYVKYVDSCEPBPOLMKNE	360		
DB 301	ROASLEISPIITFLTAQTLLMDLGOFLLSCHISSHOHDGMEAYVKYVDSCEPBPOLMKNE	360		
QY 361	EAEDYDDDLTDSMDVVRPDDNSPFTQIRSAVKGPKTWVHYIAABEEDWDYAPLVLA	420		
DB 361	EAEDYDDDLTDSMDVVRPDDNSPFTQIRSAVKGPKTWVHYIAABEEDWDYAPLVLA	420		
QY 421	DDRSYKSOYLNNGPORIGRKVKYKVRPMAYTDEPKTEATQHESGILGPLLYGVGDTL	480		
DB 421	DDRSYKSOYLNNGPORIGRKVKYKVRPMAYTDEPKTEATQHESGILGPLLYGVGDTL	480		
QY 481	LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILGCEIFKYKWTVTVEGDP	540		
DB 481	LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILGCEIFKYKWTVTVEGDP	540		
QY 541	TKSDPRCLTRYSSFFVNMRDLASGLIGPLLCYKESVDQKGNQIMSKRNVILFSVFDE	600		
DB 541	TKSDPRCLTRYSSFFVNMRDLASGLIGPLLCYKESVDQKGNQIMSKRNVILFSVFDE	600		
QY 601	NESVLTENIORFLPNPAGVQLEDPEPCASNTMHSINGVFDLSQLSVCLHEVAYWYLS	660		
DB 601	NESVLTENIORFLPNPAGVQLEDPEPCASNTMHSINGVFDLSQLSVCLHEVAYWYLS	660		
QY 661	IGAQTDFLSVFPFGYTFKHKVYEDTLTLPFPGSETVPMSENPGLWILGCHNSDFNRNG	720		
DB 661	IGAQTDFLSVFPFGYTFKHKVYEDTLTLPFPGSETVPMSENPGLWILGCHNSDFNRNG	720		
QY 721	MTALLKVSCKDNTGDYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQKNATTI	780		
DB 721	MTALLKVSCKDNTGDYEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTKQKNATTI	780		
QY 781	PENDTEKTDPMFAETPMKTCNVSSSLLMLROSPTPHGLSLSDLOEAKVETSDDPS	840		
DB 781	PENDTEKTDPMFAETPMKTCNVSSSLLMLROSPTPHGLSLSDLOEAKVETSDDPS	840		
QY 841	PGAIDSNNSLSEMTFRPQLHSGDMVTFPSGLQLRLNEXLGTATAELKKLPKVSST	900		
DB 841	PGAIDSNNSLSEMTFRPQLHSGDMVTFPSGLQLRLNEXLGTATAELKKLPKVSST	900		
QY 901	SNNLISITPSDNLAAAGTNTSSLPSPMPVYDSDLTTLFGKXSSPLTESGGPLSGEE	960		
DB 901	SNNLISITPSDNLAAAGTNTSSLPSPMPVYDSDLTTLFGKXSSPLTESGGPLSGEE	960		
QY 961	NNDKSLLESGLMNSQESSGNVNSTESGRLFKGRAGHPALLTKDNALPKVTSLLKTN	1020		
DB 961	NNDKSLLESGLMNSQESSGNVNSTESGRLFKGRAGHPALLTKDNALPKVTSLLKTN	1020		

1021 KTSNKSATNRKTHIDGPSLLIENSPSVWQNILESDETEPKKVTPLIHDRMDKNAIALRL 1080
1021 KTSNKSATNRKTHIDGPSLLIENSPSVWQNILESDETEPKKVTPLIHDRMDKNAIALRL 1080
1081 NHMSNKTTSKKNEMVQKKEGPIPPDAONPDMSPFKMLFLPESARWQRTGKNSLNSG 1140
1081 NHMSNKTTSKKNEMVQKKEGPIPPDAONPDMSPFKMLFLPESARWQRTGKNSLNSG 1140
1141 QGSPKQLVSLGPEKSVQGNFLSEKKNVVGKGFETKDVGLKEMVFPSSRNLFNTLND 1200
1141 QGSPKQLVSLGPEKSVQGNFLSEKKNVVGKGFETKDVGLKEMVFPSSRNLFNTLND 1200
1201 LHENNTNHOEKKIQEIEIEKKETLIQENVVLPOIHTVTG*KNFMKNLFLLLSTRQNVGSYD 1260
1201 LHENNTNHOEKKIQEIEIEKKETLIQENVVLPOIHTVTG*KNFMKNLFLLLSTRQNVGSYD 1260
1261 GAYAPVLODFRSLNDSTNRKKTHTAHFSKKGEEENLEGNGNOTKQIVEKYACTTISNPT 1320
1261 GAYAPVLODFRSLNDSTNRKKTHTAHFSKKGEEENLEGNGNOTKQIVEKYACTTISNPT 1320
1321 SQCNFVTOQSKKALQKFRPLPLETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE 1380
1321 SQCNFVTOQSKKALQKFRPLPLETELEKRIIVDDTSTQSKNMKHLTPSTLTQIDYNEKE 1380
1381 KGAITQSPISDCLTRSHSIPQANRSPPIAKVSSPPIRPIYLVTRVLPQDNSSHLPAAAY 1440
1381 KGAITQSPISDCLTRSHSIPQANRSPPIAKVSSPPIRPIYLVTRVLPQDNSSHLPAAAY 1440
1441 RKDQSVQESSHFLQCAKKNLISLALTLEMTGDQREVGLTSATNSVTKKVENTVLP 1500
1441 RKDQSVQESSHFLQCAKKNLISLALTLEMTGDQREVGLTSATNSVTKKVENTVLP 1500
1501 KPDLPKTSKVBLLPKVHIYQKDLFFETSNQSPGHLDVEGSLLOGTSCAIAKWEANRP 1560
1501 KPDLPKTSKVBLLPKVHIYQKDLFFETSNQSPGHLDVEGSLLOGTSCAIAKWEANRP 1560
1561 GKVPFLRVATESSAKTPSKLLDPLANDNHYGQIPKEWKSGQKSPKTAFAKKOTILSL 1620
1561 GKVPFLRVATESSAKTPSKLLDPLANDNHYGQIPKEWKSGQKSPKTAFAKKOTILSL 1620
1621 NACESHAIAAINEGONKPEIEVWAKQRTKELSCONPPVLKSHOREITRTTLQSDQEE 1680
1621 NACESHAIAAINEGONKPEIEVWAKQRTKELSCONPPVLKSHOREITRTTLQSDQEE 1680
1681 IDYDDTISYEMKKEDEDIYDEBENQSPRSFQKTRHYFIAAVERLDYGMSSSPHYLRNR 1740
1681 IDYDDTISYEMKKEDEDIYDEBENQSPRSFQKTRHYFIAAVERLDYGMSSSPHYLRNR 1740
1741 AQSGSVPOKVKVQFQFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFRQASR 1800
1741 AQSGSVPOKVKVQFQFTDGSFTQPLRYGELNEHLGLGPYIRAEVEDNIMVTFRQASR 1800
1801 PYSFYSSLISYEDDQQAEPKRNFKVKNETKTYFWKQVCHNAPTDEDFCKAWAYFSDV 1860
1801 PYSFYSSLISYEDDQQAEPKRNFKVKNETKTYFWKQVCHNAPTDEDFCKAWAYFSDV 1860
1861 DLEKDVHSGLIGLPLVCHNTNLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRA 1920
1861 DLEKDVHSGLIGLPLVCHNTNLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRA 1920
1921 PCNIQMEDTFEKNYRFAHNGVIMDTLPLGVNAQDORIRWYLLSNGSNENIHSIRPSGH 1980
1921 PCNIQMEDTFEKNYRFAHNGVIMDTLPLGVNAQDORIRWYLLSNGSNENIHSIRPSGH 1980
1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKSKAGIWRVECLIGSHLAGMSTLPLVYSNKC 2040
1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKSKAGIWRVECLIGSHLAGMSTLPLVYSNKC 2040
2041 QTPPLGNASGHIRDFQITASQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
2041 QTPPLGNASGHIRDFQITASQYQGWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
2101 HGKTKQARQKFSLSIYSQFIIMSLDGKKQTYRGNSGTGLMVFFGNVDSSGIKNIFN 2160

Db 2101 HGKTKQARQKFSLSIYSQFIIMSLDGKKQTYRGNSGTGLMVFFGNVDSSGIKNIFN 2160
Qy 2161 PPIIARYIRLPHTHYSIRSTIRLMELMGDLNCSMPLGMSKASDAQITASSYFTNFA 2220
Db 2161 PPIIARYIRLPHTHYSIRSTIRLMELMGDLNCSMPLGMSKASDAQITASSYFTNFA 2220
Qy 2221 TWSPSKARLHIQGRSNARPOVNNPKEWLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHIQGRSNARPOVNNPKEWLQVDFQKTMKVTGTTQGVKSLLTSMYVKEFL 2280
Qy 2281 ISSSDGCHQWTLFFQNGKVKVQFNQDSFTFVNSLDPLLTTRYLRHPQSWHQAIALRM 2340
Db 2281 ISSSDGCHQWTLFFQNGKVKVQFNQDSFTFVNSLDPLLTTRYLRHPQSWHQAIALRM 2340
Qy 2341 EVLGCERQDLY 2351
Db 2341 EVLGCERQDLY 2351
RESULT 8
AAW11343
ID AAW11343 standard; protein; 2351 AA.
XX AC AAW11343;
XX DT 17-NOV-1997 (first entry)
XX DE Active Factor VIII:C analogue S224X.
XX KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
XX KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
XX KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
XX XX proteolytic cleavage.
OS Homo sapiens.
OS Synthetic.
XX FH Location/Qualifiers
FT 1. .19
FT Peptide /note= "signal peptide"
FT Protein 20. .2351 /note= "mature Factor VIII:C"
FT Region 20. .1667 /note= "heavy chain fragment"
FT Modified-site 243 /label= Phe, Glu, Pro
FT Domain 760. .1667 /note= "B domain"
FT Region 1668. .2350 /note= "light chain fragment"
XX MO9703195-A1.
XX PN 30-JAN-1997.
XX PD 09-JUL-1996; 96WO-US011444.
XX PF 11-JUL-1995; 95US-0001025P.
XX PR (CHIR) CHIRON CORP.
XX PA Hung DT, Cohen FE, Innis M;
XX PI WPI; 1997-119050/11.
XX DR Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX PT used in the treatment of haemophiliacs, by improvement of haemostasis.
XX PS Claim 10; Page; 90pp; English.
XX CC AAW11330-W11472 represent active Factor VIII:C analogues of the
invention. These sequences were created by mutating the wild type Factor

VIII:C coding sequence (see AAT51357) using mutagenic primers. The analogues comprise a native Factor VIII:C polypeptide modified at a site adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg dipeptide is created. Factor VIII:C is a large glycoprotein that participates in the blood coagulation cascade that ultimately converts soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A deficiency in Factor VIII:C is responsible for haemophilia A, which is an X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is activated by plasma proteases, such as thrombin. During activation the mature polypeptide is cleaved to generate heavy and light chain fragments that are further cleaved. Complexes of two or more of the analogues, nucleic acids and vectors encoding them may be used alone or in conjunction with each other, for the prevention or treatment of active Factor VIII:C deficiency in a mammal. The analogues may be used as immunogens to raise antibodies, and in the treatment of haemophilias, by improvement of haemostasis. The analogues are resistant to proteolytic cleavage and display increased plasma half-life. They may be administered at lower dosages and by different modes of administration

XX Sequence 2351 AA;

Query March 99.8%; Score 12395; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQEILSTCFELCLARFCFSATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPSFPPN 60
DB 1 MQEILSTCFELCLARFCFSATRRYYLGAVELSWDMQSDLGELPVDARPPRPVPSFPPN 60

QY 61 TSVYVKTLFVETDHLFNIAKPRPPMGLGPTQAEVYDVTVITLKNMASHPVSLHAV 120
DB 61 TSVYVKTLFVETDHLFNIAKPRPPMGLGPTQAEVYDVTVITLKNMASHPVSLHAV 120

QY 121 GSYWKASGAEYDQTSOREKEDDKVRPGGSHYVQVLKENGMAASPLCLTYSLSH 180
DB 121 GSYWKASGAEYDQTSOREKEDDKVRPGGSHYVQVLKENGMAASPLCLTYSLSH 180

QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSHSETKNSLMQDRD 240
DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFILLFAVPDEGKSHSETKNSLMQDRD 240

QY 241 AASARAWPKMTVNGVYKNSRLPGLIGCHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNH 300
DB 241 AASARAWPKMTVNGVYKNSRLPGLIGCHRKSVYVHVGMTTPEVHSIFLEGHTFLVRNH 300

QY 301 ROASLEISPIITLTACTLMDLQGLFCHISHQHGDGMEAYVKVDSQCPREPQLMKNNE 360
DB 301 ROASLEISPIITLTACTLMDLQGLFCHISHQHGDGMEAYVKVDSQCPREPQLMKNNE 360

QY 361 EAEDYDDDLTDSEMDVVRPDDDNSPSFIQIRSAKKHPTWVHYTAAEEDWDYAPLVLA 420
DB 361 EAEDYDDDLTDSEMDVVRPDDDNSPSFIQIRSAKKHPTWVHYTAAEEDWDYAPLVLA 420

QY 421 PDRSYKSYLNGQRIQKVKYRPMAYTDETFKTREAIQHSIGILFLYGEVGDTL 480
DB 421 PDRSYKSYLNGQRIQKVKYRPMAYTDETFKTREAIQHSIGILFLYGEVGDTL 480

QY 481 LIIFKNOASRPNIYPHGIITVRLPYSLRPLKGVKHLKDFPILPGBIFKYKMTVTVEDGP 540
DB 481 LIIFKNOASRPNIYPHGIITVRLPYSLRPLKGVKHLKDFPILPGBIFKYKMTVTVEDGP 540

QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGELLICYKESVDQGNQIMSKRNVLIFSVPDE 600
DB 541 TKSDPRCLTRYSSFVNMRDLASGLIGELLICYKESVDQGNQIMSKRNVLIFSVPDE 600

QY 601 NRSWYLTENIQRLNPAGVQLEDPEFQASNMHNSINGVYFDSLQLSVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRLNPAGVQLEDPEFQASNMHNSINGVYFDSLQLSVCLHEVAYWYILS 660

QY 661 IGAQDPLSVFSGYTFKHKMYEDTLTLFPRSGETVMSMENPGLWILGCHNSDFNRG 720
DB 661 IGAQDPLSVFSGYTFKHKMYEDTLTLFPRSGETVMSMENPGLWILGCHNSDFNRG 720

QY 721 MTALLKVSQCDKNTGDYEDSDIEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVSQCDKNTGDYEDSDIEDISAYLLSKNNAIEPRSPQNSRHPSTROKQFNATTI 780

QY 781 PNDIEKTPWFAHRTMPKIQNVSSDILMLLRQSPTPHGLSLDLQEAKEYTSDPDS 840
DB 781 PNDIEKTPWFAHRTMPKIQNVSSDILMLLRQSPTPHGLSLDLQEAKEYTSDPDS 840

QY 841 PGALDNNLSLSMTHFRQLHSGDMVTPESGLQRLNEKLTGTAATLKLDFKVSST 900
DB 841 PGALDNNLSLSMTHFRQLHSGDMVTPESGLQRLNEKLTGTAATLKLDFKVSST 900

QY 901 SNMLISTPSONIAAGTNTSSLGPPSMVHVYDSQDITLFGKKSPILTESGGPISLSE 960
DB 901 SNMLISTPSONIAAGTNTSSLGPPSMVHVYDSQDITLFGKKSPILTESGGPISLSE 960

QY 961 NNDKSLLESGLMNSQESSGKNSVSTESGRLPFGKRAHGPALLTKDNALFKVISLSLKTN 1020
DB 961 NNDKSLLESGLMNSQESSGKNSVSTESGRLPFGKRAHGPALLTKDNALFKVISLSLKTN 1020

QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVMQNTILSDTEPKVTPLTHDSMLMDKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVMQNTILSDTEPKVTPLTHDSMLMDKNATALRL 1080

QY 1081 NHMSNKTSSKNMEMVQOKKGPDPDAQNPDMSFFKMLFLPESARWQRTHGKNSLNSG 1140
DB 1081 NHMSNKTSSKNMEMVQOKKGPDPDAQNPDMSFFKMLFLPESARWQRTHGKNSLNSG 1140

QY 1141 QGSPKQLVSLGPEKSVGQNFLEKKNVYVGKGFTKDVGLKEMVFPSSRNLFITNLDN 1200
DB 1141 QGSPKQLVSLGPEKSVGQNFLEKKNVYVGKGFTKDVGLKEMVFPSSRNLFITNLDN 1200

QY 1201 LHENNTHQEKKIQBEIEKKEKTLIQENVVLPOIHTVGTGKFMKNLFLSTRQNVGSD 1260
DB 1201 LHENNTHQEKKIQBEIEKKEKTLIQENVVLPOIHTVGTGKFMKNLFLSTRQNVGSD 1260

QY 1261 GAYAVLQDFRSLNDSTNRKXKTAHFSSKGBEENLEGIGNOTKQIYEKYACTTISPT 1320
DB 1261 GAYAVLQDFRSLNDSTNRKXKTAHFSSKGBEENLEGIGNOTKQIYEKYACTTISPT 1320

QY 1321 SQQNVTVQSSKRALQFRLPIEETLEKRIIVDDTSTOWSKNMKHLTSTLTQIDYNEKE 1380
DB 1321 SQQNVTVQSSKRALQFRLPIEETLEKRIIVDDTSTOWSKNMKHLTSTLTQIDYNEKE 1380

QY 1381 KGAIQTSPSLDCLTSHSIPOANRSPFLIAKVSSPESIRPIYLTVLFPQDNSSHLPAA 1440
DB 1381 KGAIQTSPSLDCLTSHSIPOANRSPFLIAKVSSPESIRPIYLTVLFPQDNSSHLPAA 1440

QY 1441 RKKDSGVQESSHFLQGAKKNNLSLAILTEMTGQDQREYGSIGTSATNSVTYKKVENTVLP 1500
DB 1441 RKKDSGVQESSHFLQGAKKNNLSLAILTEMTGQDQREYGSIGTSATNSVTYKKVENTVLP 1500

QY 1501 XDDLKPTSGKVLELIPKVIHYKDLFPPTTSNGSPGHLQVSGSLLOQTGEGAIKWEANRP 1560
DB 1501 XDDLKPTSGKVLELIPKVIHYKDLFPPTTSNGSPGHLQVSGSLLOQTGEGAIKWEANRP 1560

QY 1561 GKVPPLRVATSSAKTPSKLLDPLAWDNHYGTQIPKEBWKQKSPKPTAFKKDTILSL 1620
DB 1561 GKVPPLRVATSSAKTPSKLLDPLAWDNHYGTQIPKEBWKQKSPKPTAFKKDTILSL 1620

QY 1621 NACENHAIATAINEGONKPEIEVTWAKQRTRELCSQNPPLVKRQREBITTTLSQDQEE 1680
DB 1621 NACENHAIATAINEGONKPEIEVTWAKQRTRELCSQNPPLVKRQREBITTTLSQDQEE 1680

QY 1681 IDYDDTI SVMKKEDFIYDEENQSPRSFQKTRHYTAAVERLWDYGMSSSPVLNRR 1740
DB 1681 IDYDDTI SVMKKEDFIYDEENQSPRSFQKTRHYTAAVERLWDYGMSSSPVLNRR 1740

QY 1741 AQSGSVPOFKVVFQFTDGSFTQPLYRGEHNEHGLGPGYIRAEVEDINMVTFRNQASR 1800
DB 1741 AQSGSVPOFKVVFQFTDGSFTQPLYRGEHNEHGLGPGYIRAEVEDINMVTFRNQASR 1800

QY 1801 PYSFYSSLSIYEEDQORQGAEPKKNPKVNETKTYFWKQHHMAPTKDFCDKAWAYFSDV 1860


```

Db 1801 PYSFSSLSIYEDQOQAEPRKNFYKPNETXTYFWKVOHHMAPTKDEDFCKAWAFSDV 1860
QY 1861 DLEKDVHSLGILPLLCHTNTNTPAHGRQVTVQEFALPTTIFDEYKSWYFTENMERNCSA 1920
Db 1861 DLEKDVHSLGILPLLCHTNTNTPAHGRQVTVQEFALPTTIFDEYKSWYFTENMERNCSA 1920
QY 1921 PCNIQMEDPTFENYRFAHNGVINDTLPCLVMAQDQRIKRWYLLSGNSNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTFENYRFAHNGVINDTLPCLVMAQDQRIKRWYLLSGNSNENIHSIHFSGH 1980
QY 1981 VFTVRKKEYKXALYNLYSGVEETVEMLPKSGAGIWRVECLIGEHLAGNSTLFLVYSNKC 2040
Db 1981 VFTVRKKEYKXALYNLYSGVEETVEMLPKSGAGIWRVECLIGEHLAGNSTLFLVYSNKC 2040
QY 2041 QTPLEGNASGHIRDFQITAGQYQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
Db 2041 QTPLEGNASGHIRDFQITAGQYQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
QY 2101 HGKIKTOGARQKSSLYISOFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGIXHNIFN 2160
Db 2101 HGKIKTOGARQKSSLYISOFIIMYSLDGKKWOTYRGNSTGTLMVFFGNVDSSGIXHNIFN 2160
QY 2161 PPIIARYIRLHPHTHYSIRSLRMLMGCCLNCSMPLGMESKAISDAQITASSYPTNMPA 2220
Db 2161 PPIIARYIRLHPHTHYSIRSLRMLMGCCLNCSMPLGMESKAISDAQITASSYPTNMPA 2220
QY 2221 TWSPSKARLHLOGRSNAPPOVNNPKWILQVDFQKTMKVTVGTTQGVKSLLSIMYKKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAPPOVNNPKWILQVDFQKTMKVTVGTTQGVKSLLSIMYKKEFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVWNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVWNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
QY 2341 EVLGCEAQDLY 2351
Db 2341 EVLGCEAQDLY 2351

```

RESULT 9

AAW11435
ID AAW11435 standard; protein; 2351 AA.

AC AAW11435;

DT 20-NOV-1997 (first entry)

DE Active Factor VIII:C analogue S1311X.

XX Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
XX proteolytic cleavage.

OS Homo sapiens.
OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..19 /note= "signal peptide"

FT Protein 20..2351 /note= "mature Factor VIII:C"

FT Region 20..1667 /note= "heavy chain fragment"

FT Domain 760..1667 /note= "B domain"

FT Modified-site 1330

FT /label= Phe, Glu, Pro

FT Region 1668..2350

FT /note= "light chain fragment"

XX

XX

XX

XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR) CHIRON CORP.
XX Hung DT, Cohen FE, Innis N;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophiliacs, by improvement of haemostasis.
XX Claim 30; Page; 90pp; English.
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipetide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophiliacs, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;

Query Match 99.8%; Score 12395; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MQIELSTCFELCLIRFCFSATRRYYLGVAVELSWDYNQSDLGELPVDARPPRVKSPFN 60
Db 1 MQIELSTCFELCLIRFCFSATRRYYLGVAVELSWDYNQSDLGELPVDARPPRVKSPFN 60
QY 61 TSVVYKKTFLVEFTDHELFENIAKPRPPMGLLGPTTQAEVYDTVTITLKNMASHPVSLHAV 120
Db 61 TSVVYKKTFLVEFTDHELFENIAKPRPPMGLLGPTTQAEVYDTVTITLKNMASHPVSLHAV 120
QY 121 GVSYWKASEGAYDDQTSQREKEDDKVFPQSGSHYTVWQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSYWKASEGAYDDQTSQREKEDDKVFPQSGSHYTVWQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVDFDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFTLLFAVDFDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLLMDLQGFLLSCHSHQHDGMEAYKVKVDSCEPSPQIMKQNE 360
Db 301 ROASLEISPIITFLTAQTLLMDLQGFLLSCHSHQHDGMEAYKVKVDSCEPSPQIMKQNE 360
QY 361 EAEDYDDDLTDSMDVVRFDNDSPTQIRSVAKKHPTKVHVHTAAEEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDSMDVVRFDNDSPTQIRSVAKKHPTKVHVHTAAEEDWDYAPLVLA 420

QY 421 PDDRSYKQYLNNGPQIRGRKYKVRPMAYTDETEKTRAIQHESGILGPLYGEGDTL 480
DB 421 PDDRSYKQYLNNGPQIRGRKYKVRPMAYTDETEKTRAIQHESGILGPLYGEGDTL 480
QY 481 LIIIFXQASRPNIYFPHGITHYRPLYSRRLPGVGHLDKDFPLPOBIYKYKWTVTVEDGP 540
DB 481 LIIIFXQASRPNIYFPHGITHYRPLYSRRLPGVGHLDKDFPLPOBIYKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFNMRBRDLASGLIGPLLIYKESVDKGNQMSDKRNVLFSVDE 600
DB 541 TKSDPRCLTRYSSFNMRBRDLASGLIGPLLIYKESVDKGNQMSDKRNVLFSVDE 600
QY 601 NRSWYLTENIQRLPNPAGVQLEDEFEQASNIHMSINGVYFDSLQSLVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRLPNPAGVQLEDEFEQASNIHMSINGVYFDSLQSLVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGYTFKHMYVEDTLFLPPSGTIVFMSMENPGLWILGCHNSDFNRG 720
DB 661 IGAQTDFLSVFSGYTFKHMYVEDTLFLPPSGTIVFMSMENPGLWILGCHNSDFNRG 720
QY 721 MTALLKVSSCDKNTGDYEDSDYEDISAYLLSKNAIEPRFSQNSRHPSTROKQENATTI 780
DB 721 MTALLKVSSCDKNTGDYEDSDYEDISAYLLSKNAIEPRFSQNSRHPSTROKQENATTI 780
QY 781 PENDIEKTDPPFAHRTMPKIQNVSSSDILMLRQSPTHGLSLSDLOEAKYETESDDPS 840
DB 781 PENDIEKTDPPFAHRTMPKIQNVSSSDILMLRQSPTHGLSLSDLOEAKYETESDDPS 840
QY 841 PGADSNNSLSBWFHFRPOLHSGDMVFTPEGLOLRLEKLGTTAAATELKKLDFKVSST 900
DB 841 PGADSNNSLSBWFHFRPOLHSGDMVFTPEGLOLRLEKLGTTAAATELKKLDFKVSST 900
QY 901 SNNLSTIPSDNLAAAGTONTSSLGPPSPVHVDSQDITLFGKKSPLTESGGPLSLSEE 960
DB 901 SNNLSTIPSDNLAAAGTONTSSLGPPSPVHVDSQDITLFGKKSPLTESGGPLSLSEE 960
QY 961 NNDKLLRSGLMNSQESSWGKNVSTESGRLPKGRAGPALLTKDNALFKVSI SILKTN 1020
DB 961 NNDKLLRSGLMNSQESSWGKNVSTESGRLPKGRAGPALLTKDNALFKVSI SILKTN 1020
QY 1021 KTSNNSATNRKTHIDGSPILLIENSVMONILLESDETEPKVTPLTHDRMLMDKNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGSPILLIENSVMONILLESDETEPKVTPLTHDRMLMDKNATLRL 1080
QY 1081 NHMSKNTSSKNMNVQOKKEGIPPDACNPDMSPFEXMLFLPESARWQTHGKNSLNSG 1140
DB 1081 NHMSKNTSSKNMNVQOKKEGIPPDACNPDMSPFEXMLFLPESARWQTHGKNSLNSG 1140
QY 1141 QGPSKQIVSLGPEKSVQCNFLSEKNKVYVKGKFTKDVGLKEMVFPSSRNLFITNLND 1200
DB 1141 QGPSKQIVSLGPEKSVQCNFLSEKNKVYVKGKFTKDVGLKEMVFPSSRNLFITNLND 1200
QY 1201 LHEHNTHNQEKI QBEIEKKEFTLIQENVVLPQIHTVGTGKPMKNLFLLSSTRONVEGSVD 1260
DB 1201 LHEHNTHNQEKI QBEIEKKEFTLIQENVVLPQIHTVGTGKPMKNLFLLSSTRONVEGSVD 1260
QY 1261 GAYAVLQDFRSLNDSTNRKKTAAHPSKKGEEENLEGLGNQTKQIVEKYACTTRISPT 1320
DB 1261 GAYAVLQDFRSLNDSTNRKKTAAHPSKKGEEENLEGLGNQTKQIVEKYACTTRISPT 1320
QY 1321 SOQNVTORSGKALKQFLPLEETELEKRIIYDDTSTOWSKNMKHLTPTSLTQIDYNEKE 1380
DB 1321 SOQNVTORSGKALKQFLPLEETELEKRIIYDDTSTOWSKNMKHLTPTSLTQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRPLPAAKVSFSPSIRPIYLTIRVLQDNSSHLPAASY 1440
DB 1381 KGAITQSPSLDCLTRSHSIPQANRPLPAAKVSFSPSIRPIYLTIRVLQDNSSHLPAASY 1440
QY 1441 RKKDSGVQESSHFLQCAKKNLSAILTLEMTGDQREYVSLGTSATNSVYTKKVENTVLP 1500
DB 1441 RKKDSGVQESSHFLQCAKKNLSAILTLEMTGDQREYVSLGTSATNSVYTKKVENTVLP 1500
QY 1501 KFDLPKTSKQVLLLPKVHIYOKDLFPFTETSNQSPGHLDLVEGSLIQGTGEGAIKWEANRP 1560

DB 1501 KFDLPKTSKQVLLLPKVHIYOKDLFPFTETSNQSPGHLDLVEGSLIQGTGEGAIKWEANRP 1560
QY 1561 GKVPPLRVATSSSARTPSKLLDPLAWDNHYGTQIPKEEWKQOEKSPKTAFFKKDTILSL 1620
DB 1561 GKVPPLRVATSSSARTPSKLLDPLAWDNHYGTQIPKEEWKQOEKSPKTAFFKKDTILSL 1620
QY 1621 NACSNSHAIANAINEGONKPEIEVTWAKQRTERRCSONPPVLKGFHOREITRTTILQSDQEE 1680
DB 1621 NACSNSHAIANAINEGONKPEIEVTWAKQRTERRCSONPPVLKGFHOREITRTTILQSDQEE 1680
QY 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKTRHYPIAAVERLWDYCMSSSSPHVLNRR 1740
DB 1681 IDYDDTISVENKKEDFDIYDEDENQSPRSFQKTRHYPIAAVERLWDYCMSSSSPHVLNRR 1740
QY 1741 AQSGSVPOFKKVFQBEFTDGSFTQPLIRGELENEHLGLLGPVIRAEVEDNIMVTFNQASR 1800
DB 1741 AQSGSVPOFKKVFQBEFTDGSFTQPLIRGELENEHLGLLGPVIRAEVEDNIMVTFNQASR 1800
QY 1801 PYSFYSSLSIYEEQROGAEPKQNFVKENETKTPWKVQEHMAPTKDRFDCKAWAYFSDV 1860
DB 1801 PYSFYSSLSIYEEQROGAEPKQNFVKENETKTPWKVQEHMAPTKDRFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGILGPLLVCHTINTLNPAGHQVTVQBEFALFTTIFDETKSWYFTENMERNCRA 1920
DB 1861 DLEKDVHSLGILGPLLVCHTINTLNPAGHQVTVQBEFALFTTIFDETKSWYFTENMERNCRA 1920
QY 1921 PCNTQMEDPTKENYRPHAINGYINDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHFSGH 1980
DB 1921 PCNTQMEDPTKENYRPHAINGYINDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHFSGH 1980
QY 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPLGMAHGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
DB 2041 QTPLGMAHGHIRDFOITASGOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMII 2100
QY 2101 HGIKTQGARQFSSLIYSQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNGVDSSGIKHNIFN 2160
DB 2101 HGIKTQGARQFSSLIYSQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNGVDSSGIKHNIFN 2160
QY 2161 PPIIARYIRLRHPTFYSIRSTRMELMGCDLNSCMPLGWESKATSDAOITASSYFTNMA 2220
DB 2161 PPIIARYIRLRHPTFYSIRSTRMELMGCDLNSCMPLGWESKATSDAOITASSYFTNMA 2220
QY 2221 TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSLTTSMTVYKBEFL 2280
DB 2221 TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSLTTSMTVYKBEFL 2280
QY 2281 ISSSDQGHQWTLFFONGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPQSVWHQIALRM 2340
DB 2281 ISSSDQGHQWTLFFONGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPQSVWHQIALRM 2340
QY 2341 EVLGCFAQDLY 2351
DB 2341 EVLGCFAQDLY 2351
RESULT 10
AAW11445
ID AAW11445 standard; protein; 2351 AA.
XX
AC AAW11445;
XX
DT 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue L1643X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; anti;

KW	proteolytic cleavage.	QY	121	GVSYKASEGAEYDDQTSOREKEDDKVPFGSGSHYVWQVLKENGPMASDPCLCLTYSYLSH	180
XX		DB	121	GVSYKASEGAEYDDQTSOREKEDDKVPFGSGSHYVWQVLKENGPMASDPCLCLTYSYLSH	180
OS	Homo sapiens.	QY	181	VDLVKDLNSGLIGALLVCRSGSLAKEKTQTHKILLPAVDFDEGKSHSETKNSIMQORD	240
OS	Synthetic.	DB	181	VDLVKDLNSGLIGALLVCRSGSLAKEKTQTHKILLPAVDFDEGKSHSETKNSIMQORD	240
PH	Key	QY	241	AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYVHVIWGMGTTPVHSTFLGSHFLVNH	300
FT	Peptide	DB	241	AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYVHVIWGMGTTPVHSTFLGSHFLVNH	300
FT	Protein	QY	301	RQASLEISPIITFLTAQTLLMDLGFLLSCHISSHQHDGMEAYVKVDSCPEPQPMKNE	360
FT	Region	DB	301	RQASLEISPIITFLTAQTLLMDLGFLLSCHISSHQHDGMEAYVKVDSCPEPQPMKNE	360
FT	Domain	QY	361	EAEDYDDDLTSEMDVTRFDDDNPSFTQIRSVAKKHPKTWVHYIAAEEEDWDVAPLVIA	420
FT	Modified-site	DB	361	EAEDYDDDLTSEMDVTRFDDDNPSFTQIRSVAKKHPKTWVHYIAAEEEDWDVAPLVIA	420
FT	Label= Phe, Glu, Pro	QY	421	PDDRSYKSOYLNNGPQIRKVKVRPMAYTDETFKTREAIQHESGILGPLLYGEVDTL	480
FT	Region	DB	421	PDDRSYKSOYLNNGPQIRKVKVRPMAYTDETFKTREAIQHESGILGPLLYGEVDTL	480
PN	WO9703195-A1.	QY	481	LIIFKQASRPYNTYPHGTTDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVEGDP	540
XX	30-JAN-1997.	DB	481	LIIFKQASRPYNTYPHGTTDVRPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVEGDP	540
XX	09-JUL-1996; 95WO-US011444.	QY	541	TKSDPRCLTRYSSFFVNMERDLASGLIGPLILCYKESVDQKGNQIMSKRVILFSVDE	600
XX	11-JUL-1995; 95US-0001025P.	DB	541	TKSDPRCLTRYSSFFVNMERDLASGLIGPLILCYKESVDQKGNQIMSKRVILFSVDE	600
XX	{CHIR } CHIRON CORP.	QY	601	NRSWLTENIORFLPNPAGVLEDEPEFOASNMHSINGYVFDLSOLSVCLEHVAYWYLS	660
XX		DB	601	NRSWLTENIORFLPNPAGVLEDEPEFOASNMHSINGYVFDLSOLSVCLEHVAYWYLS	660
XX	Hung DT, Cohen FE, Innis M;	QY	661	IGAQTDFLSVFGSYTFKHKVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFNRG	720
XX	WPI; 1997-119050/11.	DB	661	IGAQTDFLSVFGSYTFKHKVYEDTLTLFPFSGETVFMENPGLMILGCHNSDFNRG	720
XX	Factor VIII:C analog modified adjacent to a non-activating Arg residue -	QY	721	MTALLKVSCKNTGDYEDSYEDISAYLLSKNNAIEPRSFQSRHSPSTQKQFNATTI	780
XX	used in the treatment of haemophiliacs, by improvement of haemostasis.	DB	721	MTALLKVSCKNTGDYEDSYEDISAYLLSKNNAIEPRSFQSRHSPSTQKQFNATTI	780
XX	Claim 32; Page; 90pp; English.	QY	781	PENDIEKTDPAFAHRTPMKTIQNVSSSLLMLLRQSPFHGSLSLDQEAKEYETSDPS	840
XX		DB	781	PENDIEKTDPAFAHRTPMKTIQNVSSSLLMLLRQSPFHGSLSLDQEAKEYETSDPS	840
CC	AAW11330-W11472 represent active Factor VIII:C analogues of the	QY	841	PGAIDSNNSLSEMTFRPOLHSHGDMVTPESGLQLRNEKLGTTAATLKLDFKVSST	900
CC	invention. These sequences were created by mutating the wild type Factor	DB	841	PGAIDSNNSLSEMTFRPOLHSHGDMVTPESGLQLRNEKLGTTAATLKLDFKVSST	900
CC	VIII:C coding sequence (see AAT51357) using mutagenic primers. The	QY	901	SNMLSTIPSDMLAAGTNTSSLGPPSMVHYDSQDLDTTLFGKSSPLTESGGPLISLEE	960
CC	analogues comprise a native Factor VIII:C polypeptide modified at a site	DB	901	SNMLSTIPSDMLAAGTNTSSLGPPSMVHYDSQDLDTTLFGKSSPLTESGGPLISLEE	960
CC	adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg	QY	961	NNDKSLLESGLMNSQESSWGKNVSTESGRLFKGRHAGPALLTKDNALFKVSISSLKTN	1020
CC	dipeptide is created. Factor VIII:C is a large glycoprotein that	DB	961	NNDKSLLESGLMNSQESSWGKNVSTESGRLFKGRHAGPALLTKDNALFKVSISSLKTN	1020
CC	participates in the blood coagulation cascade that ultimately converts	QY	1021	KTSNNASATNRKTHIDGPSLLIENSFVQNILEDSTEFKVTPLIHDRMLDKNATLRL	1080
CC	soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A	DB	1021	KTSNNASATNRKTHIDGPSLLIENSFVQNILEDSTEFKVTPLIHDRMLDKNATLRL	1080
CC	deficiency in Factor VIII:C is responsible for haemophilia A, which is an	QY	1081	NHMSKNTSSKNMENVQCKEGPIPPDAQNPDMSFPKMLFLPESARWORTHGKNSLNSG	1140
CC	X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is	DB	1081	NHMSKNTSSKNMENVQCKEGPIPPDAQNPDMSFPKMLFLPESARWORTHGKNSLNSG	1140
CC	activated by plasma proteases, such as thrombin. During activation the	QY	1141	QGFSPKQLVSLGPEKSVBQNFLEKKNVAVGKGEFTKDVGLKEMVFPSSRNLFJNLN	1200
CC	mature polypeptide is cleaved to generate heavy and light chain fragments	DB	1141	QGFSPKQLVSLGPEKSVBQNFLEKKNVAVGKGEFTKDVGLKEMVFPSSRNLFJNLN	1200
CC	that are further cleaved. Complexes of two or more of the analogues,	QY	1201	LHNNTHNQEKKIQDEIEKKTILIQENVVLQJHTVTGTQKPMKMLFLLSTRQNVGEGSYD	1260
CC	nucleic acids and vectors encoding them may be used alone or in	DB	1201	LHNNTHNQEKKIQDEIEKKTILIQENVVLQJHTVTGTQKPMKMLFLLSTRQNVGEGSYD	1260
CC	conjunction with each other, for the prevention or treatment of active	QY			
CC	Factor VIII:C deficiency in a mammal. The analogues may be used as	DB			
CC	immunogens to raise antibodies, and in the treatment of haemophiliacs, by	QY			
CC	improvement of haemostasis. The analogues are resistant to proteolytic	DB			
CC	cleavage and display increased plasma half-life. They may be administered	QY			
CC	at lower dosages and by different modes of administration	DB			
XX	Sequence 2351 AA;	QY			
XX		DB			
XX	Query Match	QY			
XX	Best Local Similarity 99.8%; Score 12394; DB 2; Length 2351;	DB			
XX	Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	QY			
XX		DB			
XX	1 XQIELSTCFPLCLRRCFATRYILGAVELSDVYMQSDLGELPVDARPPRPVPSFPFN	QY	1	XQIELSTCFPLCLRRCFATRYILGAVELSDVYMQSDLGELPVDARPPRPVPSFPFN	60
XX		DB	1	XQIELSTCFPLCLRRCFATRYILGAVELSDVYMQSDLGELPVDARPPRPVPSFPFN	60
XX	61 TSVYKKTFLVETFDHLFNIAKPRPPPMWGLGPTIOAEVYDVTWVILKNMASHPYSLHAV	QY	61	TSVYKKTFLVETFDHLFNIAKPRPPPMWGLGPTIOAEVYDVTWVILKNMASHPYSLHAV	120
XX		DB	61	TSVYKKTFLVETFDHLFNIAKPRPPPMWGLGPTIOAEVYDVTWVILKNMASHPYSLHAV	120

Db 1201 LHENNTNQEKKIQEBIEKKE*LIQENVVLPQIHVTVG*TKNFKNKLFLLLSTRQNEGSVD 1260
QY 1261 GAYAVLQDFSLNDSTNRKTKHTAHFSSKKGEEENLEGKGNQTKQIVKRYACTTRISPT 1320
Db 1261 GAYAVLQDFSLNDSTNRKTKHTAHFSSKKGEEENLEGKGNQTKQIVKRYACTTRISPT 1320
QY 1321 SQONFVTVQSKALQKFRPLPBEETLEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQONFVTVQSKALQKFRPLPBEETLEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSSPSPSIRPIYLTTRVLFDQNSSHLPAASY 1440
Db 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSSPSPSIRPIYLTTRVLFDQNSSHLPAASY 1440
QY 1441 RKDQSGVSSHFLOGAKKNLSLALTLTLEMTGDQREVGLSTGTSATNSVYKKEVNTVLP 1500
Db 1441 RKDQSGVSSHFLOGAKKNLSLALTLTLEMTGDQREVGLSTGTSATNSVYKKEVNTVLP 1500
QY 1501 KPDLPTSGKVELLPKVHIYQKDLFPPTETSNQSPGHLDVEGSLAQGTGALKWNEANRP 1560
Db 1501 KPDLPTSGKVELLPKVHIYQKDLFPPTETSNQSPGHLDVEGSLAQGTGALKWNEANRP 1560
QY 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTILSL 1620
Db 1561 GKVPFLRVATESAKTPSKLLDPLAWDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTILSL 1620
QY 1621 NACENHAAALNEGONKEIETWAKQGRTERLCSQNPVLKREOREITRTRTLOSDDOE 1680
Db 1621 NACENHAAALNEGONKEIETWAKQGRTERLCSQNPVLKREOREITRTRTLOSDDOE 1680
QY 1681 IDYDDTISVEMKKEPDFIDYDENOSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNRP 1740
Db 1681 IDYDDTISVEMKKEPDFIDYDENOSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNRP 1740
QY 1741 AQSGSVPOFKKVVOEFTDGSQTQPLRYGELNEHLGLGPIYRAEVEDNIMVTPRNQASR 1800
Db 1741 AQSGSVPOFKKVVOEFTDGSQTQPLRYGELNEHLGLGPIYRAEVEDNIMVTPRNQASR 1800
QY 1801 PYSFSSLLSYEDORQGRPEKKNFVKNETYTFWKVQHMAPTKDFDCKAWAYFSDV 1860
Db 1801 PYSFSSLLSYEDORQGRPEKKNFVKNETYTFWKVQHMAPTKDFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLGLPLVCHTNTLNPAHGRQVTVQEFALFTTIFDETCKSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLGLPLVCHTNTLNPAHGRQVTVQEFALFTTIFDETCKSWYFTENMERNCR 1920
QY 1921 PCNIQMEDPTPKENYRFAINCYIMDTLPGLWAQDORIRWYLLSMGNSNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTPKENYRFAINCYIMDTLPGLWAQDORIRWYLLSMGNSNENIHSIHFSGH 1980
QY 1981 VFTVRKKEEYKVALNLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGWSTFLVYSNKC 2040
Db 1981 VFTVRKKEEYKVALNLYPGVPEVTEMPLPSKAGIWRVECLIGEHLHAGWSTFLVYSNKC 2040
QY 2041 QTPLCWAGSHIRDFQITASQGVQWAPKLARLHYSGSINAWSTKBPFSMKIVDLLAPMII 2100
Db 2041 QTPLCWAGSHIRDFQITASQGVQWAPKLARLHYSGSINAWSTKBPFSMKIVDLLAPMII 2100
QY 2101 HGKIQGARQKFSYIYSQFIIMYSLDGKKWQTYGKSTGTLMVFPNGVDSGIGKHNFN 2160
Db 2101 HGKIQGARQKFSYIYSQFIIMYSLDGKKWQTYGKSTGTLMVFPNGVDSGIGKHNFN 2160
QY 2161 PPIIARYIRLHPTHYSIRSLTRELMLGCDLNSCPLMGSKAIDSAQITASSYFTNWEA 2220
Db 2161 PPIIARYIRLHPTHYSIRSLTRELMLGCDLNSCPLMGSKAIDSAQITASSYFTNWEA 2220
QY 2221 TWSPKARLHLQGRNAMPQVNNPREMLQVDFQKTMKVTVTTQGVKSLTSMYVXKFL 2280
Db 2221 TWSPKARLHLQGRNAMPQVNNPREMLQVDFQKTMKVTVTTQGVKSLTSMYVXKFL 2280
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTPTVWNSLDP2LLTRYLRIRHPQSWHQAIALRM 2340

Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNDSTPTVWNSLDP2LLTRYLRIRHPQSWHQAIALRM 2340
QY 2341 EVLGCEAODLY 2351
Db 2341 EVLGCEAODLY 2351
RESULT 11
AAW11398
ID AAW11398 standard; protein; 2351 AA.
XX AAW11398;
XX AC AAW11398;
XX DT 18-NOV-1997 (first entry)
XX DE Active Factor VIII:C analogue, delta 746, + residue 746 insertion.
XX KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW proteolytic cleavage.
XX CS Homo sapiens.
XX OS Synthetic.
XX FH Key
XX FT Peptide
XX FT /note= "signal peptide"
XX FT Protein
XX FT /note= "mature Factor VIII:C"
XX FT Region
XX FT /note= "heavy chain fragment"
XX FT Domain
XX FT /note= "B domain"
XX FT Misc-difference
XX FT /note= "site of 1 residue deletion"
XX FT Modified-site
XX FT /note= "inserted residue"
XX FT Region
XX FT /note= "light chain fragment"
XX PN WO9703195-A1.
XX PD 30-JAN-1997.
XX PP 09-JUL-1996; 96WO-US011444.
XX PR 11-JUL-1995; 95US-0001025P.
XX PA (CHIR) CHIRON CORP.
XX PI Hung DT, Cohen FE, Innis X;
XX DR WPI; 1997-119050/11.
XX PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophiliacs, by improvement of haemostasis.
XX PS Claim 23; Page; 90pp; English.
XX CC AAW11398-W1472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,

CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
SQ Sequence 2351 AA;

Query Match 99.8%; Score 12394; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY	1	MQIELSTCFCLLRFCFSATRRYYLGAVELSWDMQSD-GEIPVDAZFPPRPVKSPFFN	60
DB	1	MQ-ELSTCFCLLRFCFSATRRYYLGAVELSWDMQSD-GEIPVDAZFPPRPVKSPFFN	60
QY	61	TSVVYKKTLLVEFTDHLFNIAKRPFPWMLGPTIQAEVDTVITLKNWASHPVSUHAV	120
DB	61	TSVVYKKTLLVEFTDHLFNIAKRPFPWMLGPTIQAEVDTVITLKNWASHPVSUHAV	120
QY	121	GVSYWKASGAEDDQTSQREKDDKVPFGSHTYVWVLKENGPMASDPLCTYSYLISH	180
DB	121	GVSYWKASGAEDDQTSQREKDDKVPFGSHTYVWVLKENGPMASDPLCTYSYLISH	180
QY	181	VDLVKDLNSGLIGALLVCRGSLAKEXTQTLHKFILLFAVDFDEGKSWHSETKSLMODRD	240
DB	181	VDLVKDLNSGLIGALLVCRGSLAKEXTQTLHKFILLFAVDFDEGKSWHSETKSLMODRD	240
QY	241	AASARAWPKHTVNGYVNRSLPGLIGHRKSVYVHWVIGMTTPEVIGIFLEGHTFLVRNH	300
DB	241	AASARAWPKHTVNGYVNRSLPGLIGHRKSVYVHWVIGMTTPEVIGIFLEGHTFLVRNH	300
QY	301	QASLETSPIITELTAOTLLMDLGOFLLSCHISSHQHDGMEAYVWVSCPEPOLIMKXNE	360
DB	301	QASLETSPIITELTAOTLLMDLGOFLLSCHISSHQHDGMEAYVWVSCPEPOLIMKXNE	360
QY	361	EADYDDDLTDSMDVVRPDDNSPSPFIQIRSAVKHPEKTWHVYIAAEEDMDYAPLVLA	420
DB	361	EADYDDDLTDSMDVVRPDDNSPSPFIQIRSAVKHPEKTWHVYIAAEEDMDYAPLVLA	420
QY	421	PDRSRYKSOVLNNGPQIRGRKYKVRPMAYTDETFKTRBAIQHSGILGELLGEGVGDITL	480
DB	421	PDRSRYKSOVLNNGPQIRGRKYKVRPMAYTDETFKTRBAIQHSGILGELLGEGVGDITL	480
QY	481	LIIFKQASRPYNTYPEGITDVRPLYSRRLPKGVKHLKDPPILPGETFKYKWTVVEDGP	540
DB	481	LIIFKQASRPYNTYPEGITDVRPLYSRRLPKGVKHLKDPPILPGETFKYKWTVVEDGP	540
QY	541	TKSDPRCLTRYSSVFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSDKENVILFSVFDE	600
DB	541	TKSDPRCLTRYSSVFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSDKENVILFSVFDE	600
QY	601	NRSWYLTENIQRLPNPAGVQLDPEPQASINMHSINGYVDSIQLSVCLHVAAYKYLIS	660
DB	601	NRSWYLTENIQRLPNPAGVQLDPEPQASINMHSINGYVDSIQLSVCLHVAAYKYLIS	660
QY	661	IGAQTDFLSVFFSGYTFKRWVYEDTLTLPFSGETVFMSEMPNGLWILGCHNSDFNRNG	720
DB	661	IGAQTDFLSVFFSGYTFKRWVYEDTLTLPFSGETVFMSEMPNGLWILGCHNSDFNRNG	720
QY	721	YTAALLKVSXCDKNTGDIYSDYSEDISAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTI	780
DB	721	YTAALLKVSXCDKNTGDIYSDYSEDISAYLLSKNNAIEPRFSQNSRHPSTRQKQFNATTI	780
QY	781	PENDIEKTDWPAHRTPMKIQNVSSDLMLVLSQSTPHGLSLDLOEAKYTFSDDDPS	840
DB	781	PENDIEKTDWPAHRTPMKIQNVSSDLMLVLSQSTPHGLSLDLOEAKYTFSDDDPS	840
QY	841	PGAIDSNNSLSEMTFRPOLHSGDMVFTPESGQLRLNEKLGTTAATELKKLDKFKVSS	900
DB	841	PGAIDSNNSLSEMTFRPOLHSGDMVFTPESGQLRLNEKLGTTAATELKKLDKFKVSS	900

QY	901	SNNLISTIPSDNLAACTDNTSSILGPPSPVHYDSQDITLFGKKSPLTSGGPILSIEE	960
DB	901	SNNLISTIPSDNLAACTDNTSSILGPPSPVHYDSQDITLFGKKSPLTSGGPILSIEE	960
QY	961	NNDKLLSGLXNMQSSQSWGKNVSSSTESGRFLFKGRAHGPALLTKONALFKVSIILKTN	1020
DB	961	NNDKLLSGLXNMQSSQSWGKNVSSSTESGRFLFKGRAHGPALLTKONALFKVSIILKTN	1020
QY	1021	KTSNNSATNEKTHIDGPSILLIENSPPSWQNILESDETFKKVTPLIHDMMDKNAVALRL	1080
DB	1021	KTSNNSATNEKTHIDGPSILLIENSPPSWQNILESDETFKKVTPLIHDMMDKNAVALRL	1080
QY	1081	NHMSNKTTSKKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIOETHCKNSLNSG	1140
DB	1081	NHMSNKTTSKKNMEMVQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIOETHCKNSLNSG	1140
QY	1141	QGPSPKQLVSLGPEKSVGQNFLESEKXKVVGKGFETKDVGLKEMVFPSSRNLFNLND	1200
DB	1141	QGPSPKQLVSLGPEKSVGQNFLESEKXKVVGKGFETKDVGLKEMVFPSSRNLFNLND	1200
QY	1201	LHENNTNHOEKKLOEBIEKKEKTLIQENVVLPOIHTVGTGNFMKNLFLISTRQNVGSVD	1260
DB	1201	LHENNTNHOEKKLOEBIEKKEKTLIQENVVLPOIHTVGTGNFMKNLFLISTRQNVGSVD	1260
QY	1261	GAYAPVLQDFRSLNDSTNETKCHTAFFSKKGBEENLEGLGNQTKQIVKVCATTRISPT	1320
DB	1261	GAYAPVLQDFRSLNDSTNETKCHTAFFSKKGBEENLEGLGNQTKQIVKVCATTRISPT	1320
QY	1321	SOQNFVQTSKRALKQFLPLBEETELEKRIIVDDTSTQMSKMKHLPSTFLQIDYNEKE	1380
DB	1321	SOQNFVQTSKRALKQFLPLBEETELEKRIIVDDTSTQMSKMKHLPSTFLQIDYNEKE	1380
QY	1381	KGAITQSPSDCLTRSHSIPOANRSPLPIAKYSSFPSIRPIYLTVLPQDNSSHLPAASY	1440
DB	1381	KGAITQSPSDCLTRSHSIPOANRSPLPIAKYSSFPSIRPIYLTVLPQDNSSHLPAASY	1440
QY	1441	RKKGSGVQSSSHFLQAKKXNNLSLAILTLEMTGDQREVSGLTSATNSVYKKEVNTVLP	1500
DB	1441	RKKGSGVQSSSHFLQAKKXNNLSLAILTLEMTGDQREVSGLTSATNSVYKKEVNTVLP	1500
QY	1501	KDPLPTSKQVHELLPKVHIYQKDLFFTSTNSGPGHLDIVEGSLLOGTGCAIKWNEANRP	1560
DB	1501	KDPLPTSKQVHELLPKVHIYQKDLFFTSTNSGPGHLDIVEGSLLOGTGCAIKWNEANRP	1560
QY	1561	GKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEEWKSOBKSPKFAFKKDTILSL	1620
DB	1561	GKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEEWKSOBKSPKFAFKKDTILSL	1620
QY	1621	NACESNEAJAANEGQNKPEIEVTWAKQGRTERLCSQNPPLVKRHOEITRRTTLQSDQEE	1680
DB	1621	NACESNEAJAANEGQNKPEIEVTWAKQGRTERLCSQNPPLVKRHOEITRRTTLQSDQEE	1680
QY	1681	IDYDDTISVEMKKEDEDIYDEDENOSPRFOKTRHYFIAAVERLWDYGMSSSPHVLNR	1740
DB	1681	IDYDDTISVEMKKEDEDIYDEDENOSPRFOKTRHYFIAAVERLWDYGMSSSPHVLNR	1740
QY	1741	AQSGSVQPKVXVQFQFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVFNQASR	1800
DB	1741	AQSGSVQPKVXVQFQFTDGSFTQPLRYGELNEHGLLGPYIRAEVEDNIMVFNQASR	1800
QY	1801	PYSFYSSLISYEDDQROGAEPKRNFKVPNETKTYFWKHMAPTKDEDFCKAWAYFSDV	1860
DB	1801	PYSFYSSLISYEDDQROGAEPKRNFKVPNETKTYFWKHMAPTKDEDFCKAWAYFSDV	1860
QY	1861	DLEKDVHSGLIQPLLVCHVTNTINPAHGRQVTVQBFALFTIIFDETKSWYFTENMERCEA	1920
DB	1861	DLEKDVHSGLIQPLLVCHVTNTINPAHGRQVTVQBFALFTIIFDETKSWYFTENMERCEA	1920
QY	1921	PCNIQWEDTFFKENYRFFAINGYIWDITLPGVMAQORIRWYLLSNGSNENIHSIHFSGH	1980
DB	1921	PCNIQWEDTFFKENYRFFAINGYIWDITLPGVMAQORIRWYLLSNGSNENIHSIHFSGH	1980

QY 1981 VFTVRKKEEYKALNLYPGVETVEMLPFSKAGIWRVECLIGELHAGNSTLFLVYSNKC 2040
 Db 1981 VFTVRKKEEYKALNLYPGVETVEMLPFSKAGIWRVECLIGELHAGNSTLFLVYSNKC 2040
 QY 2041 QTPLGMAASHIRDFQITASQYQGWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
 Db 2041 QTPLGMAASHIRDFQITASQYQGWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
 QY 2101 HGKGTQAROKESSYISQFIIMSLDGKKWQTYRGSNGTLMVFFGNVDSGGIKENIFN 2160
 Db 2101 HGKGTQAROKESSYISQFIIMSLDGKKWQTYRGSNGTLMVFFGNVDSGGIKENIFN 2160
 QY 2161 PPIIARYIRLHPHYHSIRSLRMEIWMGDLNCSMPLGWESKAIQSDAQITASSYFTNFA 2220
 Db 2161 PPIIARYIRLHPHYHSIRSLRMEIWMGDLNCSMPLGWESKAIQSDAQITASSYFTNFA 2220
 QY 2221 TWSPSKARLHLOGRANARPOVNNPEKWLQVDFQTKMKTGVTTCGVKSLLTSMYVKEFL 2280
 Db 2221 TWSPSKARLHLOGRANARPOVNNPEKWLQVDFQTKMKTGVTTCGVKSLLTSMYVKEFL 2280
 QY 2281 ISSSDQGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPELTRYLRIHPQSWVHQIALRM 2340
 Db 2281 ISSSDQGHQWTLFFQNGKVKVFGNQDSFTPVVNSLDPELTRYLRIHPQSWVHQIALRM 2340
 QY 2341 EVLGCEAODLY 2351
 Db 2341 EVLGCEAODLY 2351

RESULT 12

AAW11425

ID AAW11425 standard; protein; 2351 AA.

XX AC AAW11425;

DT 20-NOV-1997 (first entry)

DE Active Factor VIII:C analogue T1308X.

KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;
 KW proteolytic cleavage.

OS Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FH Peptide 1..19

FT /note= "signal peptide"

FT Protein 20..2351

FT /note= "mature Factor VIII:C"

FT Region 20..1667

FT /note= "heavy chain fragment"

FT Domain 760..1667

FT /note= "B domain"

FT Modified-site 1327

FT /label= Phe, Glu, Pro

FT Region 1668..2350

FT /note= "light chain fragment"

XX WO9703195-A1.

PN 30-JAN-1997.

XX 09-JUL-1996; 96WO-US011444.

XX 11-JUL-1995; 95US-0001025P.

XX (CHIR) CHIRON CORP.

XX Hung DT, Cohen FE, Innis M;

XX

WPI; 1997-119050/11.

Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 used in the treatment of haemophilias, by improvement of haemostasis.

Claim 28; Page; 90pp; English.

AAW11330-W11472 represent active Factor VIII:C analogues of the
 invention. These sequences were created by mutating the wild type Factor
 VIII:C coding sequence (see AARS1357) using mutagenic primers. The
 analogues comprise a native Factor VIII:C polypeptide modified at a site
 adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 dipeptide is created. Factor VIII:C is a large glycoprotein that
 participates in the blood coagulation cascade that ultimately converts
 soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 activated by plasma proteases, such as thrombin. During activation the
 mature polypeptide is cleaved to generate heavy and light chain fragments
 that are further cleaved. Complexes of two or more of the analogues,
 nucleic acids and vectors encoding them may be used alone or in
 conjunction with each other, for the prevention or treatment of active
 Factor VIII:C deficiency in a mammal. The analogues may be used as
 immunogens to raise antibodies, and in the treatment of haemophilias, by
 improvement of haemostasis. The analogues are resistant to proteolytic
 cleavage and display increased plasma half-life. They may be administered
 at lower dosages and by different modes of administration

Sequence 2351 AA;

Query Match	95.8%;	Score 12394;	DB 2;	Length 2351;
Best Local Similarity	95.8%;	Pred. No. 0;		
Matches 2347;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;

QY 1 MQI8LSTCFFLCILRFFCFSSATRRYYLGAVELSWDYMQSDLGELPVDARFPRVPSFPPN 60
 Db 1 MQI8LSTCFFLCILRFFCFSSATRRYYLGAVELSWDYMQSDLGELPVDARFPRVPSFPPN 60

QY 61 TSVYVKKTLFVEFTDHLFNIAKPRPPWMLGLGPTQIAEVYTVITLKNMASHPVSLHAV 120
 Db 61 TSVYVKKTLFVEFTDHLFNIAKPRPPWMLGLGPTQIAEVYTVITLKNMASHPVSLHAV 120

QY 121 GVSYWKASEGAEYDDQTSOREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
 Db 121 GVSYWKASEGAEYDDQTSOREKEDDKVFPGGSHYVWQVLKENGPMASDPLCLTYSYLSH 180

QY 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLKFTILLFAVDEGKSWHSETKNSLMQDRD 240
 Db 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLKFTILLFAVDEGKSWHSETKNSLMQDRD 240

QY 241 AASAPAMPKMTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPVHSIFLSGHTFLVRNH 300
 Db 241 AASAPAMPKMTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPVHSIFLSGHTFLVRNH 300

QY 301 RQASLEISPIITFLTAQTLMDLQGLFLLCHLSSHQHDGMEAYVKVDSCPEBPQLMKNE 360
 Db 301 RQASLEISPIITFLTAQTLMDLQGLFLLCHLSSHQHDGMEAYVKVDSCPEBPQLMKNE 360

QY 361 BAEDYDDDLTDSEMDVVRPDDNSPFSFIQIRSAKXKHPKTVWHVYIAAEEEDWDYAPLVLA 420
 Db 361 BAEDYDDDLTDSEMDVVRPDDNSPFSFIQIRSAKXKHPKTVWHVYIAAEEEDWDYAPLVLA 420

QY 421 PDRSYKSOYLNGPQRIGRKYKVRPMAYTDEFTKTREATCHESGILGPLLYGEVGDTL 480
 Db 421 PDRSYKSOYLNGPQRIGRKYKVRPMAYTDEFTKTREATCHESGILGPLLYGEVGDTL 480

QY 481 LIIFNQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPFILPGETPKYKWTWVEDGP 540
 Db 481 LIIFNQASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPFILPGETPKYKWTWVEDGP 540

QY 541 TKSDPRCLTRYSSFFVXKVERDLASGLIGPLLCYKESVDQKGNQIMSDKRNVLIFSVPDE 600
 Db 541 TKSDPRCLTRYSSFFVXKVERDLASGLIGPLLCYKESVDQKGNQIMSDKRNVLIFSVPDE 600

QY 601 NRSWLTENIQRLPNPAGVQLEDPEPQASNMHSINGVYVFDLSQLSVCLHEVAYWYLS 660
 Db 601 NRSWLTENIQRLPNPAGVQLEDPEPQASNMHSINGVYVFDLSQLSVCLHEVAYWYLS 660
 QY 661 IGAQTDFLSVFFSGYTPFKHWVYEDTTLTPFGSETVFMENPGLWILCHNSDFNRG 720
 Db 661 IGAQTDFLSVFFSGYTPFKHWVYEDTTLTPFGSETVFMENPGLWILCHNSDFNRG 720
 QY 721 MTALLKVSSCDKNTGDYEDSYEDISAYLLSKNAIEPRFSQNSHPSTQKQFNAITI 780
 Db 721 MTALLKVSSCDKNTGDYEDSYEDISAYLLSKNAIEPRFSQNSHPSTQKQFNAITI 780
 QY 781 PENDIEKTDWFAHRTMPKIQNVSSDDLMLLRQSPTHGLSLSDLOEAKYTFSDPDS 840
 Db 781 PENDIEKTDWFAHRTMPKIQNVSSDDLMLLRQSPTHGLSLSDLOEAKYTFSDPDS 840
 QY 841 PGAIIDNSNISEMTHPRQLHHSQDMVFTPEBSGLQLRLNEKLGHTAATELKKLDKFSVST 900
 Db 841 PGAIIDNSNISEMTHPRQLHHSQDMVFTPEBSGLQLRLNEKLGHTAATELKKLDKFSVST 900
 QY 901 SNNLSTIPSDNLAAGTDNTSSLGPPSPMPVHYDSQDITLFGKSSPLTESGGPLSSEE 960
 Db 901 SNNLSTIPSDNLAAGTDNTSSLGPPSPMPVHYDSQDITLFGKSSPLTESGGPLSSEE 960
 QY 961 NNDKLLSGLMNSQSSWGKNSVSTESGRLFKGKRAHGPAALLTKONALFKVSIULKTN 1020
 Db 961 NNDKLLSGLMNSQSSWGKNSVSTESGRLFKGKRAHGPAALLTKONALFKVSIULKTN 1020
 QY 1021 KTSNNSATNKTILDGPSLLIENSPPWQNILEDSTEFKKVTLIHDRMLMDKNAALRL 1080
 Db 1021 KTSNNSATNKTILDGPSLLIENSPPWQNILEDSTEFKKVTLIHDRMLMDKNAALRL 1080
 QY 1081 NMSNKTTSKKNMVMQKKEGIPDPAQNPDXSFFKMLPLPESARWIORTHGKNSLNSG 1140
 Db 1081 NMSNKTTSKKNMVMQKKEGIPDPAQNPDXSFFKMLPLPESARWIORTHGKNSLNSG 1140
 QY 1141 QGSPKQLVSLGPEKSVGQNFISEKNKVVVGKGEFTKQVGLKEMVFPSSRNLFNLNDN 1200
 Db 1141 QGSPKQLVSLGPEKSVGQNFISEKNKVVVGKGEFTKQVGLKEMVFPSSRNLFNLNDN 1200
 QY 1201 LHNNTNHNQKKEEIEKKEKTELLIQENVLPQHTVTGKNFMKNLFLSTRQNVGSYD 1260
 Db 1201 LHNNTNHNQKKEEIEKKEKTELLIQENVLPQHTVTGKNFMKNLFLSTRQNVGSYD 1260
 QY 1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKGHEENLEGLNQTKQIVEKYACTTRISPN 1320
 Db 1261 GAYAPVLQDFRSLNDSNTRTKHTAHFSKKGHEENLEGLNQTKQIVEKYACTTRISPN 1320
 QY 1321 SQQNFVTSKRALKOPRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEXE 1380
 Db 1321 SQQNFVTSKRALKOPRLPLETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEXE 1380
 QY 1381 KGAITQPSLSDCLTRSHSIPQANRSPLPKAKVSSFPISIRPIYLTVLFDQNSHSLPAASV 1440
 Db 1381 KGAITQPSLSDCLTRSHSIPQANRSPLPKAKVSSFPISIRPIYLTVLFDQNSHSLPAASV 1440
 QY 1441 RKQDSGVQSSHPLOQAKXNNLSAILTLEMTGDQREVSLGTSATNSVTYKKEVENTVLP 1500
 Db 1441 RKQDSGVQSSHPLOQAKXNNLSAILTLEMTGDQREVSLGTSATNSVTYKKEVENTVLP 1500
 QY 1501 KPDLPTSGKVELLPKVHIYQKDLPTETSNQSGFGLDLVEGSLLOGTGAIKWNENRPP 1560
 Db 1501 KPDLPTSGKVELLPKVHIYQKDLPTETSNQSGFGLDLVEGSLLOGTGAIKWNENRPP 1560
 QY 1561 GKVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEWKSQKSPKTAFFKKOTILSL 1620
 Db 1561 GKVPFLRVATESAKTPSKLLDPLANDNHYGTQIPKEWKSQKSPKTAFFKKOTILSL 1620
 QY 1621 NACESNHAIAINEGNQKPEIEVTWAKQKTERLCSQNPVPLKRRHOREITRITLQSDQEE 1680
 Db 1621 NACESNHAIAINEGNQKPEIEVTWAKQKTERLCSQNPVPLKRRHOREITRITLQSDQEE 1680

QY 1681 IDYDDTTISVMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
 Db 1681 IDYDDTTISVMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
 QY 1741 AQSGSVPEQFKKVVFOEFTDGSFTQPLRYGELNHLGLGPIYRAEVEDNIMVTFRQASR 1800
 Db 1741 AQSGSVPEQFKKVVFOEFTDGSFTQPLRYGELNHLGLGPIYRAEVEDNIMVTFRQASR 1800
 QY 1801 PYSFYSSLISYEBDQOQGAEPKRNFKVKNETKTYFWKVQHMAPTKDEEDCKAWAYFSDV 1860
 Db 1801 PYSFYSSLISYEBDQOQGAEPKRNFKVKNETKTYFWKVQHMAPTKDEEDCKAWAYFSDV 1860
 QY 1861 DIEKDVHSLIGLLVCHNTNLPNPAHGRQVTVQEFALFTIIPDETYSWTFENMENCRA 1920
 Db 1861 DIEKDVHSLIGLLVCHNTNLPNPAHGRQVTVQEFALFTIIPDETYSWTFENMENCRA 1920
 QY 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSENENHISFSGH 1980
 Db 1921 PCNIQMEDPTFKENYRFAHNGYIMDTLPGLVMAQDQRIRWYLLSMGSENENHISFSGH 1980
 QY 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTIPLVYSNKC 2040
 Db 1981 VFTVRKEEYKMALYNLYPGVFETVEMLPKSKAGIWRVECLIGEHLHAGMSTIPLVYSNKC 2040
 QY 2041 QTPLGMAHGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
 Db 2041 QTPLGMAHGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
 QY 2101 HGKTCQARQKTSYISQFIINYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIEN 2160
 Db 2101 HGKTCQARQKTSYISQFIINYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKENIEN 2160
 QY 2161 PPIIARYIELHPTHYSIRSTLRMELMGCDLNSCSPGLMGESKAI SDAQITASSYFTNMEA 2220
 Db 2161 PPIIARYIELHPTHYSIRSTLRMELMGCDLNSCSPGLMGESKAI SDAQITASSYFTNMEA 2220
 QY 2221 TWSPSKARLHLOGRSNAWRPQVNNPKWLQVDFQKTMKVTGVTTCQVKSLLTSMVYKVEL 2280
 Db 2221 TWSPSKARLHLOGRSNAWRPQVNNPKWLQVDFQKTMKVTGVTTCQVKSLLTSMVYKVEL 2280
 QY 2281 ISSSQDGHQWTLFPQNGKVKVFGQNDSTPPVNSLDPPLLRILRIHPQSWVHQIALRM 2340
 Db 2281 ISSSQDGHQWTLFPQNGKVKVFGQNDSTPPVNSLDPPLLRILRIHPQSWVHQIALRM 2340
 QY 2341 EVLGCEAODLY 2351
 Db 2341 EVLGCEAODLY 2351
 RESULT 13
 AAW11362
 ID AAW11362 standard; protein; 2351 AA.
 XX
 AC AAW11362;
 XX
 DT 18-NOV-1997 (first entry)
 XX
 DE Active Factor VIII:C analogue L277X.
 XX
 KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
 KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
 KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
 XX proteolytic cleavage.
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Location/Qualifiers
 FT 1..19
 FT /note= "signal peptide"
 FT 20..2351
 FT /note= "mature Factor VIII:C"
 FT 20..1667
 FT Region

FT Modified-site /note= "heavy chain fragment"
 FT 296
 FT /label= pte, Glu, Pro
 FT Domain 760..1667
 FT /note= "3 domain"
 FT Region 1668..2350
 FT /note= "light chain fragment"
 XX
 PN WO9703195-A1.
 XX
 XX 30-JAN-1997.
 XX
 XX 09-JUL-1996; 96WO-US011444.
 XX
 XX 11-JUL-1995; 95US-0001025P.
 XX
 XX (CHIR) CHIRON CORP.
 XX
 XX Hung DT, Cohen FB, Innis M;
 XX WPI; 1997-119050/11.
 XX
 XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
 PT used in the treatment of haemophilias, by improvement of haemostasis.
 XX
 PS Claim 14; Page; 90pp; English.
 XX
 CC AAW11330-W11472 represent active Factor VIII:C analogues of the
 CC invention. These sequences were created by mutating the wild type Factor
 CC VIII:C coding sequence (see AAW1357) using mutagenic primers. The
 CC analogues comprise a native Factor VIII:C polypeptide modified at a site
 CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
 CC dipeptide is created. Factor VIII:C is a large glycoprotein that
 CC participates in the blood coagulation cascade that ultimately converts
 CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
 CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
 CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
 CC activated by plasma proteases, such as thrombin. During activation the
 CC mature polypeptide is cleaved to generate heavy and light chain fragments
 CC that are further cleaved. Complexes of two or more of the analogues,
 CC nucleic acids and vectors encoding them may be used alone or in
 CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration
 XX
 SQ Sequence 2351 AA;

Query Match 99.8%; Score 12394; DB 2; Length 2351;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQIELSTCFECLLPFCPSATRYVLGAVELSWDMQSDLGELPVDASPPRPVPSFPFN 60
 DB 1 MQIELSTCFECLLPFCPSATRYVLGAVELSWDMQSDLGELPVDASPPRPVPSFPFN 60
 QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWGLLGPTTQAEVYDTVTITLKNMASHPVSLHAV 120
 DB 61 TSVVYKTLFVEFTDHLFNIAKPRPPWGLLGPTTQAEVYDTVTITLKNMASHPVSLHAV 120
 QY 121 GYSYWKASEGABYDQTSQREKEDKVPFGSGHTYVWVYKENGPMASDPCLTYSYLSH 180
 DB 121 GYSYWKASEGABYDQTSQREKEDKVPFGSGHTYVWVYKENGPMASDPCLTYSYLSH 180
 QY 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFTILLFAVDEGKSWSEKNSLMQDRD 240
 DB 181 VDLVKDLSGLIGALLVCREGSLAKEKTQTLHKFTILLFAVDEGKSWSEKNSLMQDRD 240
 QY 241 AASAAWPKMHTVNGVYVNSLPLGLICGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVRNH 300
 DB 241 AASAAWPKMHTVNGVYVNSLPLGLICGCHRKSVYWHVIGMGTTPVHSIFLEGHTFLVRNH 300

QY 301 PQASLEISPIITFLTAQTLLMDLGOFLLSCHISSHOHDGMEAVVYVDSPERPOLMKNE 360
 DB 301 PQASLEISPIITFLTAQTLLMDLGOFLLSCHISSHOHDGMEAVVYVDSPERPOLMKNE 360
 QY 361 EADYDDDLTDSMDVVRFDNDSPSFTQIRSAVKGPKTWVHYIAAEEEDWDYAPLVLA 420
 DB 361 EADYDDDLTDSMDVVRFDNDSPSFTQIRSAVKGPKTWVHYIAAEEEDWDYAPLVLA 420
 QY 421 PDDRSYKSOYLNNNGPQIRGRYKVKYRPNAYTDETFKTRATQHRSGLLGPLLYGVGDTL 480
 DB 421 PDDRSYKSOYLNNNGPQIRGRYKVKYRPNAYTDETFKTRATQHRSGLLGPLLYGVGDTL 480
 QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPCEIIFKYKWTVTVEGDP 540
 DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPCEIIFKYKWTVTVEGDP 540
 QY 541 TKSDPRCLTRYYSFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSKRNVLFSVEDE 600
 DB 541 TKSDPRCLTRYYSFVNMERDLASGLIGPLLI CYKESVDQKGNQIMSKRNVLFSVEDE 600
 QY 601 NRSWYLTENIQRFLEPNPAGVQLEDEPEFOASIMHSINCYVDSLSLSVCLHEVAYWYLS 660
 DB 601 NRSWYLTENIQRFLEPNPAGVQLEDEPEFOASIMHSINCYVDSLSLSVCLHEVAYWYLS 660
 QY 661 IGAQTDFLSVFFSGYTFKHKMAYEDTLTFPFGSTVFMSMENPGLWILGCHNSDFNRG 720
 DB 661 IGAQTDFLSVFFSGYTFKHKMAYEDTLTFPFGSTVFMSMENPGLWILGCHNSDFNRG 720
 QY 721 MTALLKVSQCKNTGDDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTQKQFNATTI 780
 DB 721 MTALLKVSQCKNTGDDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTQKQFNATTI 780
 QY 781 PENDIEKTDPMFAKRTMPKIQNVSSSLLMLRQSPHPGLISLSDLOEAKYEFSDSDPS 840
 DB 781 PENDIEKTDPMFAKRTMPKIQNVSSSLLMLRQSPHPGLISLSDLOEAKYEFSDSDPS 840
 QY 841 PGADSNNSLSEMTFRPQLHSGEDMVTTPRSGIQLRLNEKLGTTAATLKKLDPKVST 900
 DB 841 PGADSNNSLSEMTFRPQLHSGEDMVTTPRSGIQLRLNEKLGTTAATLKKLDPKVST 900
 QY 901 SNLITSTPSNLAAAGTDNTSSLGPSPVHYDVSQDITLFGKSSPLTESGGPLSLSEE 960
 DB 901 SNLITSTPSNLAAAGTDNTSSLGPSPVHYDVSQDITLFGKSSPLTESGGPLSLSEE 960
 QY 961 NNDSKLLPSGLMNSQESSWGKNVSTESGRFLPKGRAGHPALLTKDNALFKVYSLLKTN 1020
 DB 961 NNDSKLLPSGLMNSQESSWGKNVSTESGRFLPKGRAGHPALLTKDNALFKVYSLLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDTEFKKVTPLIHDRMLMDKNATALRL 1080
 DB 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDTEFKKVTPLIHDRMLMDKNATALRL 1080
 QY 1081 NEMSNKTTSSKNMVMQCKEGPIPPDAQNPDMAFFPKMLFLPESARWQRTGKNSLNSG 1140
 DB 1081 NEMSNKTTSSKNMVMQCKEGPIPPDAQNPDMAFFPKMLFLPESARWQRTGKNSLNSG 1140
 QY 1141 QGPSPKQVLSLGPKEKSVGQNFLEKKNVVGKGEFTKVGLEKMPSPSNLFTNLDN 1200
 DB 1141 QGPSPKQVLSLGPKEKSVGQNFLEKKNVVGKGEFTKVGLEKMPSPSNLFTNLDN 1200
 QY 1201 LHENNTNQEKKIQEEIEKKTETLIQENVLPQIHVTGTGKNFMKNLFLLSQNVGEGSYD 1260
 DB 1201 LHENNTNQEKKIQEEIEKKTETLIQENVLPQIHVTGTGKNFMKNLFLLSQNVGEGSYD 1260
 QY 1261 GAYAPVLQDPFSLNDSNTNRKKTAAHFSKKGEENLEGLGNQTKQIVEKACTTRIGENT 1320
 DB 1261 GAYAPVLQDPFSLNDSNTNRKKTAAHFSKKGEENLEGLGNQTKQIVEKACTTRIGENT 1320
 QY 1321 SQQNFVTQSRKRALQKQRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
 DB 1321 SQQNFVTQSRKRALQKQRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380


```
QY 1381 KGAITQSPSDCITRSHSIPQANRSPPIAKVSSPFSIRPIYLTRVLFQDNGSHLPAASY 1440
Db      |||
QY 1381 KGAITQSPSDCITRSHSIPQANRSPPIAKVSSPFSIRPIYLTRVLFQDNGSHLPAASY 1440
Db      |||
QY 1441 RKDSGVQSSHPLOAKKNN:SLAILTLBMGDQREVSGLSGTSATNSVYKKVENTVLP 1500
Db      |||
QY 1441 RKDSGVQSSHPLOAKKNN:SLAILTLBMGDQREVSGLSGTSATNSVYKKVENTVLP 1500
Db      |||
QY 1501 KPDLPTSGKVELLPKVHIYQKDLFPFET:SNQSPGHLDAVEGSLQGTGEGALKNNEANRP 1560
Db      |||
QY 1501 KPDLPTSGKVELLPKVHIYQKDLFPFET:SNQSPGHLDAVEGSLQGTGEGALKNNEANRP 1560
Db      |||
QY 1561 GKVPFLRVATSSAKTPSKLLDPLAWNDHYGTQIPKEWKSKQSKSEKTAFFKDKDTILSL 1620
Db      |||
QY 1621 NACESHAIATAINEGONKBEIEVTWAKQGTERTLCSONPVLKRHOREITRRTLOSDDQE 1680
Db      |||
QY 1621 NACESHAIATAINEGONKBEIEVTWAKQGTERTLCSONPVLKRHOREITRRTLOSDDQE 1680
Db      |||
QY 1681 IDYDDTISVEMKEDDFDIYDEBENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db      |||
QY 1681 IDYDDTISVEMKEDDFDIYDEBENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db      |||
QY 1741 AQSGSVQPKVVFQFTDGSFTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFRNQASR 1800
Db      |||
QY 1801 PYSFYSSLSYEDDOQGAEPKRFNFKNETKTYFMKVOHNAPTKDEBCKAWAYFSV 1860
Db      |||
QY 1801 PYSFYSSLSYEDDOQGAEPKRFNFKNETKTYFMKVOHNAPTKDEBCKAWAYFSV 1860
Db      |||
QY 1861 DLEKDVHSLGILGLVCHTNTLNPAHGRQVTVQEPALFFTFIDETKSWYFTENMERNCA 1920
Db      |||
QY 1861 DLEKDVHSLGILGLVCHTNTLNPAHGRQVTVQEPALFFTFIDETKSWYFTENMERNCA 1920
Db      |||
QY 1921 PCNIQMEDTFKRYFHAINGIMDTLPGLVMAQQRIRWYLLSGNSNETHSIHFSGH 1980
Db      |||
QY 1921 PCNIQMEDTFKRYFHAINGIMDTLPGLVMAQQRIRWYLLSGNSNETHSIHFSGH 1980
Db      |||
QY 1981 VFTVRKKEVKALYNLYPGVRETVMPLSKAGIWEVECLIGHLHAGMSTFLVYVSNKC 2040
Db      |||
QY 1981 VFTVRKKEVKALYNLYPGVRETVMPLSKAGIWEVECLIGHLHAGMSTFLVYVSNKC 2040
Db      |||
QY 2041 QTPLGWASGHIRDFOITASQYQGWAPKLARLHYSGSINAWSTKBPFSWKVDLLAPMII 2100
Db      |||
QY 2101 HGIKTQCARQKPSLLYSQPIIMYSLDGKKWQYRGNSGTGLMVFFGNVDSSGIKENIFN 2160
Db      |||
QY 2101 HGIKTQCARQKPSLLYSQPIIMYSLDGKKWQYRGNSGTGLMVFFGNVDSSGIKENIFN 2160
Db      |||
QY 2161 PPIIARYILHPTHYIRSTLRMLMGCDLNSCMPLGMSXALSDAQITASSYFTNMA 2220
Db      |||
QY 2221 TWSPSKARLHQRSNARWPQVNNPKXELQVDFQKTKMVTGVTQGVKSLTSMYVKEFL 2280
Db      |||
QY 2281 ISSSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
Db      |||
QY 2341 EVLGCEAQDLY 2351
Db      |||
QY 2341 EVLGCEAQDLY 2351
Db      |||
```

RESULT 14

AAW11461

ID AAW11461 standard; prote.in; 2351 AA.

XX

```
AC AAW11461;
XX 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue V1717X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophiliac; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
XX Synthetic.
EH Key
FT Peptide
FT Protein
FT Region
FT Domain
FT Region
FT Modified-site
XX
XX WO9703195-A1.
XX 30-JAN-1997.
XX 09-JUL-1996; 96WO-US011444.
XX 11-JUL-1995; 95US-0001025P.
XX (CHIR ) CHIRON CORP.
XX Hung DT, Cohen PE, Innis M;
XX WPI; 1997-119050/11.
XX Factor VIII:C analog modified adjacent to a non-activating Arg residue -
XX used in the treatment of haemophiliacs, by improvement of haemostasis.
XX Claim 36; Page; 90pp; English.
XX
XX AAW11330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAT51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophiliacs, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;
```

Query Watch

Best Local Similarity 99.8%; Score 12394; DB 2; Length 2351;

Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

1 MOELSTCFCLLRFCFSATRYVYLGAVELSDYMOQSDLGELPVDARPPBRVPKSPFN 60
Db 1 MOELSTCFCLLRFCFSATRYVYLGAVELSDYMOQSDLGELPVDARPPBRVPKSPFN 60
QY 61 TSVVYKKTFLVEFTDHLFNIAKRPMPMGLGPTIOAEVYDVTVITLKNASHFVSLHAV 120
Db 61 TSVVYKKTFLVEFTDHLFNIAKRPMPMGLGPTIOAEVYDVTVITLKNASHFVSLHAV 120
QY 121 GYSYWKASGABYDDOTSOREKEDDXYFEGGSHTYUWQVLKENGPNASDPLCLITYSLSH 180
Db 121 GYSYWKASGABYDDOTSOREKEDDXYFEGGSHTYUWQVLKENGPNASDPLCLITYSLSH 180
QY 181 VDLVKDLSGLICALLVCEGSLAKEKTOILHFIILLFAVDFGKSWHSETKNSLQDRED 240
Db 181 VDLVKDLSGLICALLVCEGSLAKEKTOILHFIILLFAVDFGKSWHSETKNSLQDRED 240
QY 241 AASARAWPKMHTVNGYVNRSLPGLIGHRKSIVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVNGYVNRSLPGLIGHRKSIVYWHVIGMGTTPPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISITITFLTAOTLLMDLQGLLSCHISSEHQDGMAYVYKVDSCPEEPOLIKWNE 360
Db 301 ROASLEISITITFLTAOTLLMDLQGLLSCHISSEHQDGMAYVYKVDSCPEEPOLIKWNE 360
QY 361 EABDYDDDLTDSBMDVVRFDNDSPSFIQIRSVAKKHPKTWHYIAAEBEDWDYAPLVLA 420
Db 361 EABDYDDDLTDSBMDVVRFDNDSPSFIQIRSVAKKHPKTWHYIAAEBEDWDYAPLVLA 420
QY 421 PDRSRYKSYLNGPQRIQRYKVKFMAVTDTPKTREAIQHESGILGPLLYGEVGDTL 480
Db 421 PDRSRYKSYLNGPQRIQRYKVKFMAVTDTPKTREAIQHESGILGPLLYGEVGDTL 480
QY 481 LIIFKQASRPYNIYPHGITDVRPLXSRRLPKGVKHLKOPPLPGIFKYKXTVTVEDGP 540
Db 481 LIIFKQASRPYNIYPHGITDVRPLXSRRLPKGVKHLKOPPLPGIFKYKXTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFYNNMERDLASGLIGPLLI CYKESVDQGNQIMSKDNVILFSVDE 600
Db 541 TKSDPRCLTRYSSFYNNMERDLASGLIGPLLI CYKESVDQGNQIMSKDNVILFSVDE 600
QY 601 NRSWYLITENIQRELNPAGVQLEDPFQASINIMESINGYVDFSLOQLSVCLHEVAYWYILS 660
Db 601 NRSWYLITENIQRELNPAGVQLEDPFQASINIMESINGYVDFSLOQLSVCLHEVAYWYILS 660
QY 661 ICAQTDPLSVFFSGYTFKHQWYEDTLTPFPFSGEYVFMENPGLWILGCHNSDFRNRG 720
Db 661 ICAQTDPLSVFFSGYTFKHQWYEDTLTPFPFSGEYVFMENPGLWILGCHNSDFRNRG 720
QY 721 MTALLKVSSCDKNTGYEDSYEDISAYLLSKNNA-EPRSFQNSRHPSTROKQFNATTI 780
Db 721 MTALLKVSSCDKNTGYEDSYEDISAYLLSKNNA-EPRSFQNSRHPSTROKQFNATTI 780
QY 781 PENDIEKTDPLWFARHTPMKIQNVSSDILMLRQSTPHGLSLDLOEAKYETFSDDPS 840
Db 781 PENDIEKTDPLWFARHTPMKIQNVSSDILMLRQSTPHGLSLDLOEAKYETFSDDPS 840
QY 841 PGAIIDNNLSLSEMTFRPQLHSGDMVFTPEBSGLQLRLNEKLGTTAATLKKLDFKVSST 900
Db 841 PGAIIDNNLSLSEMTFRPQLHSGDMVFTPEBSGLQLRLNEKLGTTAATLKKLDFKVSST 900
QY 901 SNNLI STIPSDNLAAGTNTSSIGPSPMPVHYDSQDITLFGKKSGLTESGGPLSLSEE 960
Db 901 SNNLI STIPSDNLAAGTNTSSIGPSPMPVHYDSQDITLFGKKSGLTESGGPLSLSEE 960
QY 961 NNDKSLLESGLMNSQSSWGKNVSSYTESGELFKGKBAHGPAITKONALFKVISILKTN 1020
Db 961 NNDKSLLESGLMNSQSSWGKNVSSYTESGELFKGKBAHGPAITKONALFKVISILKTN 1020
QY 1021 KTSNNATNRKTHIDGPSLLIENSPSVWQVILSDTEFKKVTPLIHDRMLMDKNATALL 1080
Db 1021 KTSNNATNRKTHIDGPSLLIENSPSVWQVILSDTEFKKVTPLIHDRMLMDKNATALL 1080

1081 NEMSNKTTSSKNWEMVQOKKEGPIPPDAONPDMSPFPMFLPESAKWIORTHEKSLNSG 1140
Db 1081 NEMSNKTTSSKNWEMVQOKKEGPIPPDAONPDMSPFPMFLPESAKWIORTHEKSLNSG 1140
QY 1141 QGSPKQLVSLGPEKSVBQGNFLSBKNKVVGKBEFTKDVGLKEMVFPSSRLFLTNLND 1200
Db 1141 QGSPKQLVSLGPEKSVBQGNFLSBKNKVVGKBEFTKDVGLKEMVFPSSRLFLTNLND 1200
QY 1201 LHENNTHNOEKIKOBEIEKKEETLIOENVVLPOIHVITGTFKNFMKQLFLLSTQNVEGSD 1260
Db 1201 LHENNTHNOEKIKOBEIEKKEETLIOENVVLPOIHVITGTFKNFMKQLFLLSTQNVEGSD 1260
QY 1261 GAYAPVLQDFRSLNDSTNETKHTAHFSKKGEEENLEGLGNOTKQIVKYACTTRISPN 1320
Db 1261 GAYAPVLQDFRSLNDSTNETKHTAHFSKKGEEENLEGLGNOTKQIVKYACTTRISPN 1320
QY 1321 SQONFVTOSSKALQKFRLEETLEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQONFVTOSSKALQKFRLEETLEKRIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGAITOSPLSDCLTRSHSTPOANRSPPLIAKYSSPSPRIPIYLVTRVLPQDNSSHLPAA 1440
Db 1381 KGAITOSPLSDCLTRSHSTPOANRSPPLIAKYSSPSPRIPIYLVTRVLPQDNSSHLPAA 1440
QY 1441 RKDQSGVQSSSHFLQAKAKNNLSLAILTLEMTGDQREVSGSLGTSATNSVTYKVENTVLP 1500
Db 1441 RKDQSGVQSSSHFLQAKAKNNLSLAILTLEMTGDQREVSGSLGTSATNSVTYKVENTVLP 1500
QY 1501 KPDLPTSQKVBLLPKVHIYQKDLPTTSTNSGPHGLDLVEGSLQGTGGAIKWNEANRP 1560
Db 1501 KPDLPTSQKVBLLPKVHIYQKDLPTTSTNSGPHGLDLVEGSLQGTGGAIKWNEANRP 1560
QY 1561 GKVPFLVATATESAKTPSKLLDPLANDNHYGTQIKPEEMKSOEKSEKTAFFKKOTILSL 1620
Db 1561 GKVPFLVATATESAKTPSKLLDPLANDNHYGTQIKPEEMKSOEKSEKTAFFKKOTILSL 1620
QY 1621 NACESNHAAIANEGONKPEIEVWAKQGTRELCSQNPVLKRRHOREITRITLQSDQEE 1680
Db 1621 NACESNHAAIANEGONKPEIEVWAKQGTRELCSQNPVLKRRHOREITRITLQSDQEE 1680
QY 1681 IDYDDTISVEMKKE3FDIYDEBENOSPRFQKTRHYFFAAVERLMDYQMSSPHVLNRN 1740
Db 1681 IDYDDTISVEMKKE3FDIYDEBENOSPRFQKTRHYFFAAVERLMDYQMSSPHVLNRN 1740
QY 1741 AOSGSPQPKVVFQFTDGSFTQPLRGELNEHGLLGPYIRAEVEDNIMVTFRQASR 1800
Db 1741 AOSGSPQPKVVFQFTDGSFTQPLRGELNEHGLLGPYIRAEVEDNIMVTFRQASR 1800
QY 1801 PYSFYSSLISYBEDQRQGAERPNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDV 1860
Db 1801 PYSFYSSLISYBEDQRQGAERPNFVKPNETKTYFWKVQHMAPTKDEDFCKAWAYFSDV 1860
QY 1861 DLEKDVHSGLIQPLLVCHTNTLNPAHGRQVTVQEFALPTIIFDETKSWYFTENMERCSA 1920
Db 1861 DLEKDVHSGLIQPLLVCHTNTLNPAHGRQVTVQEFALPTIIFDETKSWYFTENMERCSA 1920
QY 1921 PCNIQWEDTFFKBNRFFHAINGYIMDTLPLVNAODORIRWYLLSMGSENENHSIFPSGH 1980
Db 1921 PCNIQWEDTFFKBNRFFHAINGYIMDTLPLVNAODORIRWYLLSMGSENENHSIFPSGH 1980
QY 1981 VFTVRKKEBYKMALXNLYPGVFETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVSNKC 2040
Db 1981 VFTVRKKEBYKMALXNLYPGVFETVEMLPFSKAGIWRVECLIGEHLHAGMSTLFLVSNKC 2040
QY 2041 QTPFLMGASGHIDFQITASGOYGOWAPKLARLHYSGSINAWSTKEPPSWIKVDLLAPMII 2100
Db 2041 QTPFLMGASGHIDFQITASGOYGOWAPKLARLHYSGSINAWSTKEPPSWIKVDLLAPMII 2100
QY 2101 HGIKQAPKQKSSLYISQFIIMYSIDGKKWOTYRGNSTGTLMVFFGNVSSGIKHNIFN 2160
Db 2101 HGIKQAPKQKSSLYISQFIIMYSIDGKKWOTYRGNSTGTLMVFFGNVSSGIKHNIFN 2160
QY 2161 PPIIARYIRLHPTHYSIRSTLRELMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMPA 2220

Db 2161 PPIIARYISLPHYSIRSTLRLMELMGCDLNSCMPLGWSKASDAQITASSYFTNMA 2220
QY 2221 TWSPSKARLHLOGRSNAWRPQVNNPKEWLQVDFQKTMKVTGTTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRSNAWRPQVNNPKEWLQVDFQKTMKVTGTTQGVKSLTSMYVKEFL 2280
QY 2281 TSSSQDCHQWTLFFQNGKVKVFGQNGDSFTPVVNSLDPPLTRYLRHQPQSWHQIALSM 2340
Db 2281 TSSSQDCHQWTLFFQNGKVKVFGQNGDSFTPVVNSLDPPLTRYLRHQPQSWHQIALSM 2340
QY 2341 EVLGCSAQDLY 2351
Db 2341 EVLGCSAQDLY 2351

RESULT 15
AAW11387
ID AAW11387 standard; protein; 2351 AA.
XX
AC AAW11387;
DT 18-NOV-1997 (first entry)
DE Active Factor VIII:C analogue N357X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..19
FT Protein /note= "signal peptide"
FT /note= "mature Factor VIII:C"
FT Region 20..1667
FT /note= "heavy chain fragment"
FT Modified-site 3%
FT /label= Phe, Glu, Pro
FT Domain 760..1667
FT /note= "B domain"
FT Region 1668..2350
FT /note= "light chain fragment"
XX
FN WO9703195-A1.
XX
PD 30-JAN-1997.
XX
PP 09-JUL-1996; 96WO-US011444.
XX
PR 11-JUL-1995; 95US-0001025P.
XX
PA (CHIR) CHIRON CORP.
XX
XX Hung DT, Cohen FE, Innis M;
XX WPI; 1997-119050/11.
DR
XX
PT Factor VIII:C analog modified adjacent to a non-activating Arg residue -
PT used in the treatment of haemophiliacs, by improvement of haemostasis.
XX
XX
XX Claim 20; Page; 90pp; English.
XX
CC AAW11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT51357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that

CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophiliacs, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration
XX
XX Sequence 2351 AA;
SQ
Query Match 99.8%; Score 12394; DB 2; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MQEILSTCFCLLRFCPSATRRYVIGAVELSWDYMQSDLGELPVDARPPRPVPSFBN 60
Db 1 MQEILSTCFCLLRFCPSATRRYVIGAVELSWDYMQSDLGELPVDARPPRPVPSFBN 60
QY 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLLGPPTQAEVYDTWVTLKNMASHPYSLHAV 120
Db 61 TSVVYKTLFVEFTDHLFNIAKPRPPWMLLGPPTQAEVYDTWVTLKNMASHPYSLHAV 120
QY 121 GVSVMKASGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLH 180
Db 121 GVSVMKASGAEYDDQTSQREKDDKVPFGSGHTYVWQVLKENGPMASDPLCLTYSYLH 180
QY 181 VDLVKDLNSGLICALLVCREGLAKETQTLHKFTLLFAVDECKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLICALLVCREGLAKETQTLHKFTLLFAVDECKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGVNRSLFGLICHRKSVYHVI GMGTTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPKMHTVNGVNRSLFGLICHRKSVYHVI GMGTTPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELTAQTLLMDLGQFLLSCHISSHOHGMEAYKVVDSPPEPQLMKNE 360
Db 301 ROASLEISPIITELTAQTLLMDLGQFLLSCHISSHOHGMEAYKVVDSPPEPQLMKNE 360
QY 361 EAEDYDDDLTDSEMDVWRFDNDSPSFIQIRSVAKKPKTWVHYTAAEBEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDSEMDVWRFDNDSPSFIQIRSVAKKPKTWVHYTAAEBEDWDYAPLVLA 420
QY 421 PDCRSYKCYLNNGPQRIGRKYKVRFMAYTDETFKTRAIQHESGILGPLLYGEVGDTL 480
Db 421 PDCRSYKCYLNNGPQRIGRKYKVRFMAYTDETFKTRAIQHESGILGPLLYGEVGDTL 480
QY 481 LIIFKNQASRPNTYPHGTDVRLPYSRRLPKGVKHLKDPFPLPGEIIFYKKTWTVEDGP 540
Db 481 LIIFKNQASRPNTYPHGTDVRLPYSRRLPKGVKHLKDPFPLPGEIIFYKKTWTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLICYKESVDQKGNMCKRNVLISVDFE 600
Db 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLICYKESVDQKGNMCKRNVLISVDFE 600
QY 601 NRSWYLTENIQRLFPNPAQVQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
Db 601 NRSWYLTENIQRLFPNPAQVQLEDEPQASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGTTFKHMYVETDITLTPPSGETVFMNMENPGLWILGCNDFRNG 720
Db 661 IGAQTDFLSVFSGTTFKHMYVETDITLTPPSGETVFMNMENPGLWILGCNDFRNG 720
QY 721 MTALLKVVSSCDKNTGDYVSDSDISAYLLSKNNAIEPRSPNSRHPSTROKQFNATTI 780
Db 721 MTALLKVVSSCDKNTGDYVSDSDISAYLLSKNNAIEPRSPNSRHPSTROKQFNATTI 780

QY 781 PENDIEKTDWFAHRTPEPKIQNVSSDLMRLRQSPHGLSLSDLOEAKYETESDDPS 840
DB 781 PENDIEKTDWFAHRTPEPKIQNVSSDLMRLRQSPHGLSLSDLOEAKYETESDDPS 840
QY 841 PGADSNNSLSMTFRPOLHSGDMVFTPEBSGLQRLNEKLGTTAAATELKKLDPKVSST 900
DB 841 PGADSNNSLSMTFRPOLHSGDMVFTPEBSGLQRLNEKLGTTAAATELKKLDPKVSST 900
QY 901 SNNLSTIPSDNLAAGTNTSISLGGPMPVHVDSQDITLFGKKSPLTESGGPLSLSEE 960
DB 901 SNNLSTIPSDNLAAGTNTSISLGGPMPVHVDSQDITLFGKKSPLTESGGPLSLSEE 960
QY 961 NNDKSLLESGLNCSGSKVSVSTESGRLEPKGRAHQZPALLTKDNALFKVSI SLLKTN 1020
DB 961 NNDKSLLESGLNCSGSKVSVSTESGRLEPKGRAHQZPALLTKDNALFKVSI SLLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENS SVWQNIIESDETEPKVTPLIHDMMLDKNTAURL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENS SVWQNIIESDETEPKVTPLIHDMMLDKNTAURL 1080
QY 1081 NMSNKTSSKNMVMVOQKKGPIPPDAQNPDMSPFKMLFLPESARWIORTHGKSLNSG 1140
DB 1081 NMSNKTSSKNMVMVOQKKGPIPPDAQNPDMSPFKMLFLPESARWIORTHGKSLNSG 1140
QY 1141 QGPKQLVSLGPEKSVESQNFLEBKXKVVGKGEFTKDVGLKENVFPSSRNLFNLN 1200
DB 1141 QGPKQLVSLGPEKSVESQNFLEBKXKVVGKGEFTKDVGLKENVFPSSRNLFNLN 1200
QY 1201 LHENNTNHOEKIIOBEIEKKEITLIGNVVLPOIHTVTGTQNMKNLFLSTRONVEGSD 1260
DB 1201 LHENNTNHOEKIIOBEIEKKEITLIGNVVLPOIHTVTGTQNMKNLFLSTRONVEGSD 1260
QY 1261 GAYAVLQDPRSLNDSTNRKTHAFPSKKGEENLEGLGNQTKOIVEKYACTTRISNT 1320
DB 1261 GAYAVLQDPRSLNDSTNRKTHAFPSKKGEENLEGLGNQTKOIVEKYACTTRISNT 1320
QY 1321 SQONFVTRSKRALQKPLBEETELEKRIIVDDTSTQWSKNMKLTPTSTLQIDYNEKE 1380
DB 1321 SQONFVTRSKRALQKPLBEETELEKRIIVDDTSTQWSKNMKLTPTSTLQIDYNEKE 1380
QY 1381 KGATQSPSLDCLTRSHSTPOANRPLPIAKVSSPSPRIPIYLTIVLQDNSSHPAASY 1440
DB 1381 KGATQSPSLDCLTRSHSTPOANRPLPIAKVSSPSPRIPIYLTIVLQDNSSHPAASY 1440
QY 1441 RKDSGVQESSHFLQAKKNLISAILTLEMTGDQREVGLSATNSVTYKKVENTVLP 1500
DB 1441 RKDSGVQESSHFLQAKKNLISAILTLEMTGDQREVGLSATNSVTYKKVENTVLP 1500
QY 1501 KPDLPKTSKVELLPKVHIYOKDLPTETNSGSPGHLDIVEGSLIQGTGEGAIKWNENRP 1560
DB 1501 KPDLPKTSKVELLPKVHIYOKDLPTETNSGSPGHLDIVEGSLIQGTGEGAIKWNENRP 1560
QY 1561 GKVPFLRVATESAKTPSKLDPLDLDNHYGQIPKEWKSQESKPEKTAFFKKDTILSL 1620
DB 1561 GKVPFLRVATESAKTPSKLDPLDLDNHYGQIPKEWKSQESKPEKTAFFKKDTILSL 1620
QY 1621 NACSNHAIAAINEGQNKPEIIVTWAKQRTERLCSQNPVLRKHQBRIITRTLOSQOEE 1680
DB 1621 NACSNHAIAAINEGQNKPEIIVTWAKQRTERLCSQNPVLRKHQBRIITRTLOSQOEE 1680
QY 1681 IDYDDTISVEMKKEDPDYDEBENQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNR 1740
DB 1681 IDYDDTISVEMKKEDPDYDEBENQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNR 1740
QY 1741 AQSGSVPOFKVVFQEFDTGSPQLYRGELNEHJGLLGPYIRAEVEDNIMVTFRNQASR 1800
DB 1741 AQSGSVPOFKVVFQEFDTGSPQLYRGELNEHJGLLGPYIRAEVEDNIMVTFRNQASR 1800
QY 1801 PYSFYSSLIISYEDORQGAEPKFNFKVNETKTYTWKQHHMAPTKDBFDCKAWAYFSDV 1860
DB 1801 PYSFYSSLIISYEDORQGAEPKFNFKVNETKTYTWKQHHMAPTKDBFDCKAWAYFSDV 1860
QY 1861 DLEKDVHSGSLGPLLVCNTINPAHGRQVTVQBFALFTTFDETKSWYFTENMERNCR 1920

DB 1861 DLEKDVHSGSLGPLLVCNTINPAHGRQVTVQBFALFTTFDETKSWYFTENMERNCR 1920
QY 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSMGSNENIHSIHPSGH 1980
DB 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPLGLVMAQDQIRWYLLSMGSNENIHSIHPSGH 1980
QY 1981 VFTVRKKEEYKMAIYNLYPGVFETVEMLPKAGIWEVECLIGEHLAGMSTLFLVYSNKC 2040
DB 1981 VFTVRKKEEYKMAIYNLYPGVFETVEMLPKAGIWEVECLIGEHLAGMSTLFLVYSNKC 2040
QY 2041 QTPPLGMAVSGHIRDFOITASGOYQWAPKLAIRLHYSGSINAMSTKEPFSWIKVDLIAPMII 2100
DB 2041 QTPPLGMAVSGHIRDFOITASGOYQWAPKLAIRLHYSGSINAMSTKEPFSWIKVDLIAPMII 2100
QY 2101 HGIKTQGAROKFSELYISQFIIMYSLDOKKQWTFRGNSTGTILMVFFGNVDSGIGHNIFN 2160
DB 2101 HGIKTQGAROKFSELYISQFIIMYSLDOKKQWTFRGNSTGTILMVFFGNVDSGIGHNIFN 2160
QY 2161 PPIIARYIRLHPTFYSIRSTLRMEIMGCDLNSCSMPLGWSKALISDAOITASSYFTNMFA 2220
DB 2161 PPIIARYIRLHPTFYSIRSTLRMEIMGCDLNSCSMPLGWSKALISDAOITASSYFTNMFA 2220
QY 2221 TWSPSKARLHLQGRSNARPPVNNPKEMLYQVDFQKTMKVTGVTTQGVKSLLTSMVYKEFL 2280
DB 2221 TWSPSKARLHLQGRSNARPPVNNPKEMLYQVDFQKTMKVTGVTTQGVKSLLTSMVYKEFL 2280
QY 2281 ISSSQDGHQWTLFPQNGKVKVFOGQNDSTFPVNSLDPPILTRYLRHHPQSWHQAIALRM 2340
DB 2281 ISSSQDGHQWTLFPQNGKVKVFOGQNDSTFPVNSLDPPILTRYLRHHPQSWHQAIALRM 2340
QY 2341 EVLGCEAOADLY 2351
DB 2341 EVLGCEAOADLY 2351

Search completed: April 13, 2004, 14:04:16

Job time : 77.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 27 seconds
(without alignments)
8375.788 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQIELSTCFPLCLLRFCPSA.....WVHQIALRMEVLGCEAQQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12399	99.9	2351	1 EZHU	coagulation factor
2	8320	71.0	2319	2 A47004	coagulation factor
3	8057	64.9	2133	2 T42763	coagulation factor
4	2824.5	22.8	2224	1 KFRU05	coagulation factor
5	2771.5	22.3	2183	2 T42764	coagulation factor
6	2743	22.1	2211	1 KFR005	coagulation factor
7	2190	17.6	869	2 A25945	coagulation factor
8	1326	10.7	1069	1 KUHU	ferroxidase [EC 1.11.1.10]
9	1277	10.3	1059	1 A35210	ferroxidase [EC 1.11.1.10]
10	1104	8.9	216	2 A44258	factor VIII-associ
11	663	5.3	427	2 JC4915	ags protein precu
12	657	5.3	463	1 A35479	milk fat globule m
13	650	5.2	409	2 T11743	pp47 protein - pig
14	635	5.1	401	2 S65138	Glycoprotein anti
15	635	5.1	427	2 S74211	PAS-6/7 protein pr
16	443	3.6	927	1 JQ0948	A5 antigen precurs
17	424.5	3.4	218	2 A47285	milk fat globule p
18	306.5	2.5	3133	2 S52093	hemocytin - silkw
19	262	2.1	845	2 JC5256	adipocyte transcri
20	239	1.9	1072	2 A86827	hypothetical prote
21	236.5	1.8	3418	1 G03234	breast cancer tumo
22	235.5	1.3	1358	2 A29360	SIRA protein - yea
23	227	1.8	2954	2 T14156	kinesin-related pr
24	224	1.8	2166	2 G70163	hypothetical prote
25	221.5	1.8	3329	2 T30904	breast cancer tumo
26	221	1.8	719	2 S51739	transcription reg
27	219.5	1.8	928	2 S46773	myosin heavy chain
28	219.5	1.8	3329	2 T42205	breast cancer susc
29	219	1.8	2401	2 T28676	rhopty protein -

30 218.5 1.8 3507 2 T34513 hypothetical prote
31 217.5 1.8 1271 2 D64237 hypothetical prote
32 217 1.7 1381 1 S45731 probable calcium-b
33 216.5 1.7 3328 2 T30835 breast cancer tumo
34 216 1.7 3924 2 S37431 ankryrin 2, neurona
35 213.5 1.7 2346 2 T13829 Tpr homolog - frui
36 213 1.7 1420 1 A44361 amiloride-sensitiv
37 212.5 1.7 5005 2 P82884 hypothetical prote
38 212 1.7 1957 2 T38077 hypothetical coile
39 211 1.7 1283 2 T13799 neurexin IV - frui
40 209.5 1.7 1177 2 T64233 hypothetical prote
41 208.5 1.7 1628 2 E30538 hypothetical prote
42 207.5 1.7 737 2 T13149 hypothetical prote
43 207 1.7 6669 2 S55024 nebulin, skeletal
44 206.5 1.7 737 2 T15615 hypothetical prote
45 204.5 1.6 1805 2 T02712 similar to late em

RESULT 1
EZHU
coagulation factor VIII precursor [validated] - human
N:Alternate names: antihemophilic factor A; coagulation factor VIIIc; procoagulant compon
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence revision 28-Aug-1985 #text change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445; B42:
R:Gitschier, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93265012; PMID:1303178
A:Accession: I54318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M88648; NID:gl82381; PIDN:AAA52420.1; PID:gl82383
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschier, J.; Keyt, B.; Seeburg
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman, D.D.; J
S, D.N.; Hewick, R.W.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1361-2351 <RE2>
A:Cross-references: GB:X01740; NID:gl82802; PIDN:AAA52484.1; PID:gl82803
R:Truett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; Kuo,
B.; Randolph, A.; Urdea, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen, J.; Noi
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C and the
A:Reference number: A23584; MUID:86081164; PMID:3935400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:gl82817; PIDN:AAA52485.1; PID:gl82818
R:Eaton, D.; Rodriguez, H.; Vehar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavages
ity.
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36; 392-399, 'X', 401-402; 1668-1678; 1709-1722, 'D', 1723-1725; 1741-1755 <EAT>
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

Biochemistry 31, 3315-3325, 1992
A;Title: Identification and functional importance of tyrosine sulfate residues within re
A;Reference number: A42348; MUID:92207952; PMID:1554716
A;Accession: A42348
A;Molecule type: protein
A;Residues: 20-36,356-371;392-408;582-594;1668-1669,'X',1671;1672-1692;1693-1708;1709-17
A;Experimental source: recombinant material from Chinese hamster ovary cells
A;Note: sequence extracted from NCBI backbone and corrected to correspond with the publi
R;Pay, P.J.; Smudzian, T.M.
J. Biol. Chem. 264, 14005-14010, 1989
A;Title: Intersubunit fluorescence energy transfer in human factor VIII.
A;Reference number: A43986; MUID:89340500; PMID:2503509
A;Accession: A43986
A;Molecule type: protein
A;Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
R;Leyte, A.; van Schijndel, H.B.; Niehrs, C.; Ruttnier, W.B.; Verbeet, M.P.; Marters, K.;
J. Biol. Chem. 266, 740-746, 1991
A;Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
A;Reference number: A56109; MUID:91093266; PMID:1898735
A;Contents: annotation; introns
R;McMullen, B.A.; Fujikawa, K.; Davie, E.W.; Hedner, U.; Ezban, M.
Protein Sci. 4, 740-746, 1995
A;Title: Locations of disulfide bonds and free cysteines in the heavy and light chains o
A;Reference number: A56216; MUID:95338127; PMID:7613471
A;Contents: annotation; disulfide bonds
A;Note: 329-Cys, 711-Cys, and 2019-Cys were shown to have free sulfhydryls
R;Kjalker, M.; Hedner, U.; Talbo, G.; Persson, E.; Thomsen, C.; Ezbhan, M.
Eur. J. Biochem. 234, 773-779, 1995
A;Title: Amino acid residues 721-729 are required for full factor VIII activity.
A;Reference number: S63527; MUID:96163459; PMID:8575434
A;Accession: S63527
A;Molecule type: protein
A;Residues: 733-752;753-759 <KJA>
R;bind, P.; Larsson, K.; Spira, J.; Sydow-Baackman, M.; Aïmstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A;Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A;Reference number: S66445; MUID:96048024; PMID:7556150
A;Accession: S66445
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1668-1685 <LIN>
C;Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
C;Genetics:
A;Gene: GDB:F8C
A;Cross-references: GDB:119124; OMIM:306700
A;Map position: Xg28-Xc28
A;Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
C;Function:
A;Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
A;Pathway: blood coagulation
C;Superfamily: coagulation factor VIII; discoicin I amino-terminal homology; ferroxidase
C;Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-2351/Product: coagulation factor VIII #status experimental <MAT>
F;20-740/Product: coagulation factor VIIa heavy chain #status experimental <ACH>
F;20-356/Domain: A1 <DA1>
F;23-348/Domain: ferroxidase repeat homology <F01>
F;23-759/Domain: A2 <DA2>
F;392-759/Domain: A3 <DA3>
F;402-730/Domain: ferroxidase repeat homology <F02>
F;760-1667/Domain: B <DB0>
F;1668-2351/Product: coagulation factor VIIa light chain #status experimental <ACL>
F;1709-2038/Domain: A3 <DA3>
F;1716-2038/Domain: ferroxidase repeat homology <F03>
F;2039-2191/Domain: C1 <DC1>
F;2039-2168/Domain: discoicin I amino-terminal homology <DN1>
F;2192-2351/Domain: C2 <DC2>
F;2192-2345/Domain: discoicin I amino-terminal homology <DN2>
F;60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1

F;172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status
F;355-356/cleavage site: Arg-Met (coagulation factor Xa, prorein C) #status predicted
F;355,737,738,742,1683,1698/Binding site: sulfate (Tyr) (covalent) #status experimental
F;351-382/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F;414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
F;759-760/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F;1667-1668/cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F;1708-1709/cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F;1740-1741/cleavage site: Arg-Ala (coagulation factor Xa, thrombin) #status experimental
F;2193-2345/Disulfide bonds: #status predicted
Query Match 99.9%; Score 12399; DB 1; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MQIELSTCFPLCLLRFCFSATRIYYIGAVELSDYMQSDIGELPVDARFPRPVKSPFPN 60
Db 1 MQIELSTCFPLCLLRFCFSATRIYYIGAVELSDYMQSDIGELPVDARFPRPVKSPFPN 60
QY 61 TSVYKKTLEFVEFTDHLNFAKRPMPWMLGTLGTTIAEVYDTVTITLKNMASHPVSLHAV 120
Db 61 TSVYKKTLEFVEFTDHLNFAKRPMPWMLGTLGTTIAEVYDTVTITLKNMASHPVSLHAV 120
QY 121 GVSYWKASGAEYDDQTSQREKDDKVPFGSGSHYVQVLKENGPMASDPLCLITYSLSH 180
Db 121 GVSYWKASGAEYDDQTSQREKDDKVPFGSGSHYVQVLKENGPMASDPLCLITYSLSH 180
QY 181 VDLVKDLNSGLICALLVCREGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
Db 181 VDLVKDLNSGLICALLVCREGSLAKETQTLHKFILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVGVNLSPLGLICHRKSVVHVGMTTPEVHSIFLEGTFLVRNH 300
Db 241 AASARAWPKMHTVGVNLSPLGLICHRKSVVHVGMTTPEVHSIFLEGTFLVRNH 300
QY 301 ROASLEISPTITLTAOTLMDLGQFLLSCHISSHQHDGMEAYKVDSCPEEPQLMKNE 360
Db 301 ROASLEISPTITLTAOTLMDLGQFLLSCHISSHQHDGMEAYKVDSCPEEPQLMKNE 360
QY 361 EADYDDDDTDSMDVVRDDNSPSFIOIRSVAKKPKTKWHYIAAEEDWDYAPLVLA 420
Db 361 EADYDDDDTDSMDVVRDDNSPSFIOIRSVAKKPKTKWHYIAAEEDWDYAPLVLA 420
QY 421 PDRSYKSYLNNGPQIRGRKYKVRFMAYTDTFTKTRAIQHSGLIPLLYGEVGDIL 480
Db 421 PDRSYKSYLNNGPQIRGRKYKVRFMAYTDTFTKTRAIQHSGLIPLLYGEVGDIL 480
QY 481 LIIFKQASRPNIYPHGTDVRLPYSRRLPKGVKHLKOPPLPGEIFKYKWTVTVEDGP 540
Db 481 LIIFKQASRPNIYPHGTDVRLPYSRRLPKGVKHLKOPPLPGEIFKYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMBERDLASGLIGPLLCYKESVDKGNQMSDKENVTLESVDE 600
Db 541 TKSDPRCLTRYSSFVNMBERDLASGLIGPLLCYKESVDKGNQMSDKENVTLESVDE 600
QY 601 NRSWLTENIQRLPNPAGVQLEDEPFQASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
Db 601 NRSWLTENIQRLPNPAGVQLEDEPFQASNMHSINGYVFDLSQLSVCLHEVAYWYILS 660
QY 661 IGAQTDPLSVFSGYTFKHKMYEDTLTLPSPSGTVMMSNPGLWILGCHNSDFRNRG 720
Db 661 IGAQTDPLSVFSGYTFKHKMYEDTLTLPSPSGTVMMSNPGLWILGCHNSDFRNRG 720
QY 721 MTALLKVVSSCDKNTGDYVSDYSYEDISAYLLSKNNAIEPRSPNSRHPSTROKQNAITI 780
Db 721 MTALLKVVSSCDKNTGDYVSDYSYEDISAYLLSKNNAIEPRSPNSRHPSTROKQNAITI 780
QY 781 PENDIEKTPWFARHTPMPKIQNVSSSDLLMLRQSPTPHGLSLDLOBAKYETSDDPS 840
Db 781 PENDIEKTPWFARHTPMPKIQNVSSSDLLMLRQSPTPHGLSLDLOBAKYETSDDPS 840
QY 841 PGADISNLSMTHFRQLHSHGDMVTFPSGLQRLNEKLTGTTAATLKLDPKVSST 900
Db 841 PGADISNLSMTHFRQLHSHGDMVTFPSGLQRLNEKLTGTTAATLKLDPKVSST 900

```
Db      841  PGALDSNNLSSEMTFRPQLPHSGDMVFTPESSGLQLRLNEKLGTTAAATELKKCLDFKVSST  900
Qy      901  SNMLISITPSDNLAAGTNTSSIGPPSPVHYDSOLDTTLFGKSSPLTSSGGPLSLSEE  960
Db      901  SNMLISITPSDNLAAGTNTSSIGPPSPVHYDSOLDTTLFGKSSPLTSSGGPLSLSEE  960
Qy      961  NNDSPKLESGLMNSQESSWGKNSSTESGRLPFGKGAHGPALLTKONALFKVSI SLLKTN  1020
Db      961  NNDSPKLESGLMNSQESSWGKNSSTESGRLPFGKGAHGPALLTKONALFKVSI SLLKTN  1020
Qy     1021  KTSNNSTNKTHTIDGSLIENSPPSWONILSDTEFKKVTPLIHDRMLMDKNATLRL  1080
Db     1021  KTSNNSTNKTHTIDGSLIENSPPSWONILSDTEFKKVTPLIHDRMLMDKNATLRL  1080
Qy     1081  NMSNKTTSNNKMEMVOOKKEGPTTPDAQNPDMSPFFKMLFLPESAKWIQETHGKNSLSNG  1140
Db     1081  NMSNKTTSNNKMEMVOOKKEGPTTPDAQNPDMSPFFKMLFLPESAKWIQETHGKNSLSNG  1140
Qy     1141  QGSPKQLVSLGPEKSVEGQNF-SEKNKVVVGKEFTKQVGLKEMVFPSSRNFLFLNLDN  1200
Db     1141  QGSPKQLVSLGPEKSVEGQNF-SEKNKVVVGKEFTKQVGLKEMVFPSSRNFLFLNLDN  1200
Qy     1201  LHENNTNOKKTOEIEKKEKTELIOENVLPQIHTVTGKNFMKNLFLLESTRQNVGSGYD  1260
Db     1201  LHENNTNOKKTOEIEKKEKTELIOENVLPQIHTVTGKNFMKNLFLLESTRQNVGSGYD  1260
Qy     1261  GAYAPVLQDPRSLNDSTNRKKTHTAFPSKKGEBENLEGNGNQTKQIVEKYACTRISPNT  1320
Db     1261  GAYAPVLQDPRSLNDSTNRKKTHTAFPSKKGEBENLEGNGNQTKQIVEKYACTRISPNT  1320
Qy     1321  SQCNFTQSKRALKQFRLPLEETELEKRIIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE  1380
Db     1321  SQCNFTQSKRALKQFRLPLEETELEKRIIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE  1380
Qy     1381  KGAITQSPJSDCLTRGSHSIPQANRSLPTAKVSSPFSIRPIYITRVLFDQNSHSLPAASY  1440
Db     1381  KGAITQSPJSDCLTRGSHSIPQANRSLPTAKVSSPFSIRPIYITRVLFDQNSHSLPAASY  1440
Qy     1441  RKKDSGVQSPSSHPLOQAKKNNLSLAILTLEMTGDQREVSLGTSATNSVTYKKAVENTVLP  1500
Db     1441  RKKDSGVQSPSSHPLOQAKKNNLSLAILTLEMTGDQREVSLGTSATNSVTYKKAVENTVLP  1500
Qy     1501  KPDLPKTSKRVKLLPKVHIYQKDLPTTETSNQSPGHLDLVEGSLQGTGEGAIKWNENRP  1560
Db     1501  KPDLPKTSKRVKLLPKVHIYQKDLPTTETSNQSPGHLDLVEGSLQGTGEGAIKWNENRP  1560
Qy     1561  GXVPPFLRVATESAKTPSKLLDPLANDHYGTQIPKEENKQESPEKTAFFKKDTILSL  1620
Db     1561  GXVPPFLRVATESAKTPSKLLDPLANDHYGTQIPKEENKQESPEKTAFFKKDTILSL  1620
Qy     1621  NACESNFAIAINEGONKEPPIEYTWAKQGRTERLCSQNPPLVKRHOREITRITLQSDQEE  1680
Db     1621  NACESNFAIAINEGONKEPPIEYTWAKQGRTERLCSQNPPLVKRHOREITRITLQSDQEE  1680
Qy     1681  IDYDDTISVEMKEDPDIDYDENQSPRQFKTRHYPTAAVERLMDYGNSSPHVLRNR  1740
Db     1681  IDYDDTISVEMKEDPDIDYDENQSPRQFKTRHYPTAAVERLMDYGNSSPHVLRNR  1740
Qy     1741  AQSGSVQPKVKVYQFPTDGSFTQPLVYRGELNEHLLGPLYIRAEVEDNIMVTFRQASR  1800
Db     1741  AQSGSVQPKVKVYQFPTDGSFTQPLVYRGELNEHLLGPLYIRAEVEDNIMVTFRQASR  1800
Qy     1801  PYSFYSSLISYBEDQQAEPKRNFKVNETKTYFWKVQHMAPTKDEDFCKAWAYFSDV  1860
Db     1801  PYSFYSSLISYBEDQQAEPKRNFKVNETKTYFWKVQHMAPTKDEDFCKAWAYFSDV  1860
Qy     1861  DLEKDVHSLIGPLLVCHTNTLNPAHGRQVTVQEPALFTTIDETKSWYFTENMERNCA  1920
Db     1861  DLEKDVHSLIGPLLVCHTNTLNPAHGRQVTVQEPALFTTIDETKSWYFTENMERNCA  1920
Qy     1921  PCNIQWEDPTFKENREHAINGYIMDTLPLGMAQDQRIRWYLLSGNSNENHSIHFSGH  1980
Db     1921  PCNIQWEDPTFKENREHAINGYIMDTLPLGMAQDQRIRWYLLSGNSNENHSIHFSGH  1980
```

```
Qy      1981  VFTVRKEEYKMALYNLYPCGVFTVEMLPKAGINWVECLIGBHLHAGMTTLFLVYSNKC  2040
Db      1981  VFTVRKEEYKMALYNLYPCGVFTVEMLPKAGINWVECLIGBHLHAGMTTLFLVYSNKC  2040
Qy     2041  QTEPLGMASGHIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPFSWTKVDLLAPMII  2100
Db     2041  QTEPLGMASGHIRDFQITASGOYQOWAPKLARLHYSGSINAWSTKEPFSWTKVDLLAPMII  2100
Qy     2101  HGIKTQCARQKFSLSYISQFIIMYSLDGKKWQYRGNSTGTLMWFFGNVDSSGIKENIFN  2160
Db     2101  HGIKTQCARQKFSLSYISQFIIMYSLDGKKWQYRGNSTGTLMWFFGNVDSSGIKENIFN  2160
Qy     2161  PPIIARYIRLHPETHYSIRSTLRMELMGCDLNSCMLPGMESKAI SDAQTASSYFTNMEA  2220
Db     2161  PPIIARYIRLHPETHYSIRSTLRMELMGCDLNSCMLPGMESKAI SDAQTASSYFTNMEA  2220
Qy     2221  TWSPSKARLHLOQRSNAWRPQVNNPKEMIQVDPQKTMKVGTGTVQGVKSLLTSWYKEFL  2280
Db     2221  TWSPSKARLHLOQRSNAWRPQVNNPKEMIQVDPQKTMKVGTGTVQGVKSLLTSWYKEFL  2280
Qy     2281  ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLTRYLRIHPQSWVHQIALRM  2340
Db     2281  ISSSQDGHQWTLFFQNGKVKVFGQNDSTFPVNSLDPPLTRYLRIHPQSWVHQIALRM  2340
Qy     2341  EVLGCEAODLY 2351
Db     2341  EVLGCEAODLY 2351
```

RESULT 2

A47004

coagulation factor VIII precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #-ext_change 18-Jun-1999

C:Accession: A47004

R:Ellder, B.; Lakich, D.; Girschier, J.

Genomics 16, 374-379, 1993

A:Title: Sequence of the murine factor VIII cDNA.

A:Reference number: A47004; MUID:93300511; PMID:8314577

A:Accession: A47004

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2319 <EID>

A:Cross-references: GB:105573; NID:g192456; PIDN:AAA37385.1; PID:g192457

C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase F;1-19/Domain: signal sequence #status predicted <SIG>

F;23-349/Domain: ferroxidase repeat homology <FO1>

F;402-730/Domain: ferroxidase repeat homology <FO2>

F;1686-2006/Domain: ferroxidase repeat homology <FO3>

F;2007-2156/Domain: discolidin I amino-terminal homology <DN1>

F;2160-2313/Domain: discolidin I amino-terminal homology <DN2>

Query Match

Best Local Similarity 71.0%; Score 8820; DB 2; Length 2319;

Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;

Qy 1 MQIELSTCFELCLLRPCPSATRRYVILGAVELSWDYMQSD-LGELPVDARPPRPVKSPFF 59

Db 1 MQIALFACFFLSLNFPCSSAIRRYVILGAVELSNWYIQSDLLSVLHTDSRFLPRMSTSPFF 60

Qy 60 NTSVVYKTKTLFVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVITLKNWASHPSVLSHA 119

Db 61 NTSIMYKTVFYVYKQDLFNIAKPPPMWGLLGPITWTEVHTDVTITLKNWASHPSVLSHA 120

Qy 120 VGVSYWKAEGAYDDQTSQREKDDKVPFGSGSHYVWQVLKENGPMASDPLCLITYSLYS 179

Db 121 VGVSYWKAEGAYDDQTSQREKDDKVPFGSHYVWQVLKENGPMASDPLCLITYSLYS 180

Qy 180 HYDLVKDNLNSGLIGALLVCRRGSLAKEKTOQLHKPILFAVFDGKSWHSTKNSLMQDR 239

Db 181 HYDLVKDNLNSGLIGALLVCRRGSLAKEKTOQLHKPILFAVFDGKSWHSTKNSLMQDR 240

T42763
coagulation factor VIII precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42763
R:Ullrich, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z22269
A:Status: preliminary; translated from G3/EMBL/DBJ
A:Accession: T42763
A:Molecule type: mRNA
A:Residues: 1-2133 <LOL>
A:Cross-references: EMBL:U49517; NID:G1511633; PID:G1511634; PIDN:AAB06705.1
C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2133/Product: coagulation factor VIII #status predicted <MAT>
F:23-349/Domain: ferroxidase repeat homology <FOX1>
F:402-730/Domain: ferroxidase repeat homology <FOX2>
F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match 64.9%; Score 8057; DB 2; Length 2133;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;

QY 1 MQELSTCFLLRPFCAIRRYLCAVELSWDMQSD-IGELPVDAREPFRVPSFPF 59
Db 1 MQELSTCFVCLLPLGRFAIRRYLCAVELSWDRQSELRLHVDTPFATAPALPL 60

QY 60 NTSVYKTLFVEFDHFNIAKPPPMGLLPTIAQAEVTVVITLKNMASHPVSLHA 119
Db 61 GPSVLYKTVFVEFDQLPSVARPPPMGLLPTIAQAEVTVVITLKNMASHPVSLHA 120

QY 120 VGVSYWKASEGAEYDDTSQRKEDDKVPFGSHYVWQVLZENGPMASDPCLTYSYLS 179
Db 121 VGVSYWKASEGAEYEDHTSQREKEDDKVLPKGSQTVVWQVLZENGPTASDPCLTYSYLS 180

QY 180 HYDLVKDLNSGLIGALLVCREGLAKETQTLHKPILLFAVDEGKSWHSEKNSLMQDR 239
Db 181 HYDLVKDLNSGLIGALLVCREGLTPERTQNLHFFVLLFAVDEGKSWHSAANDSWTRAM 240

QY 240 DAASARAWPMHTVNGYVNRSLPGLIGCHRKVYHVIMGTTPVHSLFEGHTFLVRN 299
Db 241 DPAPARACPMHTVNGYVNRSLPGLIGCHRKVYHVIMGTTPVHSLFEGHTFLVRN 300

QY 300 HQASLEISPIFLTAQTLMDLGOFLLSCHTSSSHQDGEAYVXVDSCEPEPOLIMKN 359
Db 301 HQASLEISPIFLTAQTLMDLGOFLLSCHTSSSHQDGEAYVXVDSCEPEPOLIMKN 360

QY 360 EBAEDYDDLTDSEMDVVRFDSDNSPFIQIRSVAKKPKTWVHYIAAEEEDWDYAPVL 419
Db 361 EB-EDYDNLVSDMDVVRFDSDNSPFIQIRSVAKKPKTWVHYIAAEEEDWDYAPVL 419

QY 420 APDRSYKSOYLNNGPQIRGRYKVRNAYTDTFTKTRTAAQHSGLGPLLYGEVGT 479
Db 420 SPDSRSYSLYNSGQIRGRYKVRNAYTDTFTKTRTAAQHSGLGPLLYGEVGT 479

QY 480 LT-IFKQASRPNTYVPHGITDVRPLYSRLPKGVKHLKDFPLAGEIFKYKWTTVEDG 539
Db 480 LLIIFKQASRPNTYVPHGITDVRPLYSRLPKGVKHLKDFPLAGEIFKYKWTTVEDG 539

QY 540 PKSDPRCLTRYSSFVNMERDLASGLIGPLLIICYKESVDQKGNQIMSDKRVILFSVFD 599
Db 540 PKSDPRCLTRYSSINLEKDLASGLIGPLLIICYKESVDQKGNQIMSDKRVILFSVFD 599

QY 600 ENRSYLTENIRFLPNPAGVQLDEPEQASNMHMSINGYVDSIQLSVCLHEVAYWYL 659
Db 600 ENQSYLAENIRFLPNPAGVQLDEPEQASNMHMSINGYVDSIQLSVCLHEVAYWYL 659

QY 660 SICAQTDLSVFPFGYTPKHKVYVEDTLTLFPFGSETFMSENGLWILGCHNSDFNR 719
Db 660 SVGAQTDLSVFPFGYTPKHKVYVEDTLTLFPFGSETFMSENGLWILGCHNSDFNR 719

QY 720 GNTALLKYSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSPQNSRHHSTKQKQNMAT 779
Db 720 GNTALLKYSCDRDIGDYNDTYEDIPGFLLSGKVIIEPRSPAQRPPSAQKQQTIT 779

QY 780 IPENDIEKTDPMFAIRTPMKIQNVSSSDLLMLLRQSPHPGLSLDLOEAKYETSDDP 839
Db 780 SPEDDVE-LDPOSGETQALELSVPSGDGSMLLQNPAPPHSGSSSDLOEARN--ADY 836

QY 840 SPGAIDSNLSSEMTHFRPQLHSGDMVFTPSG+QLRLNEKLGTTAAATLKLKLPKVS 899
Db 837 LPGAERNTAPSAARLRPELHSAERV-TPEP-----EK-----ELKLSKMS 882

QY 900 TSNWLLS--TISDMLAAGTNTSSIGPSPMPVHDSQLDITLPGKSSPLTBSGGPLSL 957
Db 883 SDLLKTSPTISDLSABTERHSLGPPHPQWFRSGLGALVGLKNSSHFTGAGVPLGS 942

QY 958 SEENDSKLLESGLMKNQBSMGKVVSTESGRLEFKGAHGPALLTKDNALFKVYSILL 1017
Db 943 TEEDH-----SSIGENVSPVESGIFEKRAHGPASLTKDDVLFKNISLV 989

QY 1018 KTKNTSNNSATNRKTHIDGPSLLIENSPSWONILESDETFKKVTPELIHDRMLDKNATA 1077
Db 990 KTKARVVLKTKRKHIDDAALLTENASA-----TFMDKNTTA 1028

QY 1078 LELNHSNKTSSKNWNVQKKGCPIPDQNPDMSEFKMLFLPESAKWQRTGKNSL 1137
Db 1029 SGLNHSN-----WIKGLGKNPL 1047

QY 1138 NSGQPSPKQVSLGPEKSVGQNFSLKKNVVGKGEFTKDVGLKEMVFPSSRLFTLN 1197
Db 1048 SSERGSPPELLTSSGSGSVKQSGQGRIRVAVSEELSG--KEMMLPNSLFTLN 1104

QY 1198 LDNLHNNTHQEKIKOBEIEKKTLLQENVVLPOIHTVGTGNFMKNLFLLSRQNVG 1257
Db 1105 SADVQNDTHSQCKSREEMERREKLVQEKVDLPVYATGATGKFLRNIFHQSTPSV 1164

QY 1258 SVDGAYAPVLQDFPSLNDSTNRTKHTAHFKSGKEENLEGKNOTKQIVEXACTRIS 1317
Db 1155 FDGGHAPVPOSESDNSAERAETHIAHFAIRBEAPLEAFGNET----- 1210

QY 1318 PNTSQNFVTORESKALKQPLFLPEETELEKRIIVDDTSTOWSKNKHLPSTLTQIDYN 1377
Db 1211 -GPGPSAVPRVKQSLQIRLPLEBIEIKPGRGVNLNSTRWS----- 1252

QY 1378 EKEGAIQTSPSLDCLTRSHSIPQANRSLPIAKVSSPSPRIYILTRVLPODNSHLPA 1437
Db 1253 ----- 1252

QY 1438 ASYRKDSDGVQESHFLOGAKKNLSLAILTLEMTGDQREYVSGLSGTSATNSVYKKVENT 1497
Db 1253 -----ESSFLOGAKKNLSLFLILEVAGGKGKISALGKSAAGPLASGKLEKA 1301

QY 1498 VLPKPDLPKTKGKVELLPKVHIYOKDLFPPTTSNGSPGHLDLVEGSLLGQTEGAKWNEA 1557
Db 1302 VLSAGLSEASGKGFPLPKVAVHREDLLPQKTSNVSCAHDGLQBI FQKTRGPNLVNKV 1361

QY 1558 NPQKVPPLRVATESSAKTPSKLDPLAWDNYGTQIPKEWKSQEKSPKTAFFKDTI 1617
Db 1362 NRPG-----RTPSKLLGP-----PMPK-EWESLEKSPKSTALTAKDII 1398

QY 1618 -LSLNACESNEALAINEGQNKPIEVTWAKQRTERLCSQNPVVLKQHOREITFTTQS 1676
Db 1399 SLPLDRHESNHSIAKNEGQATOREAANTQGGPGLCAPKPPVLRHQRDII SLPTQOP 1458

QY 1677 DOERIDYDDT-SVEMKKEDFDIYDDEENQSPRSQKTRHYFIAAVERLWDYGMSSSPHV 1736
Db 1459 EEDKMDYDDIETSTKGEDFDIYGEDENQDPRSFKQKTRHYFIAAVERLWDYGMSSSPHV 1518

QY 1737 LRNRAQSGSVQPKVVPQETDGSFTQPLVRGELNEHLGLLGYIRAEVDNIMVTRN 1796
Db 1519 LRNRAQGEVFRFKVVPREFADGSFTQPSYRGEINLHGLLGYIRAEVDNIMVTRN 1578

QY 1797 QASRPYSYSSLIIVEEDQROGASPRKNFVXPNETKITVFWKVQHMMATKDEFFCKAWAY 1856

Db 1579 QASRPYSFSSLSYSPDDQSGAEPRHNFQPNETRIYFWKVQHHMPTDEDDCKAWAY 1638
QY 1857 FSDVDLEKOVHSGVGLIGLLNCHNTLNPAHQGRQVTVQEFALPTTIDETKSWFTENNER 1916
Db 1639 FSDVDLEKOVHSGVGLIGLLNCRANTLNAAHQGRQVTVQEFALPTTIDETKSWFTENNER 1698
QY 1917 NCRAPCNQIOMEDPTFKENYRFFHAINGYIMDTPLGVNAQDQRIWYLLSGMSNENIHSIH 1976
Db 1699 NCRAPCHLQMEDPTLKENYRFFHAINGYIMDTPLGVNAQDQRIWYLLSGMSNENIHSIH 1759
QY 1977 FSGHVTVRKKEBKVALYNLYFGVFTVEMLPKSGAGIWRVECLIGEHLAGMSTLFLVY 2036
Db 1759 FSGHVSVRKKEBKCAVYNLYFGVFTVEMLPKSGAGIWRVECLIGEHLAGMSTLFLVY 1818
QY 2037 SNKQCPPLGASGHIRDFQITAGQVQWAPKARLHYSGSINAWSTKEPFSWIKVDLLA 2096
Db 1819 SKCQAPLGVASGRIRDFQITAGQVQWAPKARLHYSGSINAWSTKDPHFWIKVDLLA 1878
QY 2097 PMIHGKTQGARQKTSLSYISQFIIMYSLDGKKWOTYRGNSTGTLWVFFGNVDSGKIH 2156
Db 1879 PMIHGKTQGARQKTSLSYISQFIIMYSLDGKNWQSYRGNSTGTLWVFFGNVDSGKIH 1938
QY 2157 NFNPPILAIYIRLHTHYSIRTLRMELMGCDLNSCMPLGWESKAISDAQTASSYFT 2216
Db 1939 NFNPPIVARIYIRLHTHYSIRTLRMELMGCDLNSCMPLGMQNKAISSDQITASSHLS 1998
QY 2217 NFNATWSPSKARLHLOGRSNAWEPQNNPKWLOVDFQKTKMTYGTQGVKSLLSMYV 2276
Db 1999 NFNATWSPQARLHLOGRTNAMPVRSSABEWQVQLQKTKVGTQGVKSLLSMYV 2058
QY 2277 KPFLISSDGHQHTLFFQNGKVKVQGNQDSTPVPVNSLDPPLLYRIIRIIPQSMVHUI 2336
Db 2059 KPFLVSSQDGRWTLFLQDGHKTVFQGNQDSTPVPVNALDPLFLYRIIRIIPQSMVHUI 2118
QY 2337 ALRMEVLGCAQDLY 2351
Db 2119 ALRLEVLCGAQDLY 2133

RESULT 4

KPHUS

N/Alternate names: coagulation [validated] - human

C/Species: Homo sapiens (man)

C/Date: 19-May-1989 #sequence revision 02-Jun-1995 #text_change 08-Dec-2000

C/Accession: A56172; A42344; A28028; A27498; A25897

R/Cripe, L.D.; Moore, K.D.; Kane, W.H.

Biochemistry 31, 3777-3785, 1992

A/Title: Structure of the gene for human coagulation factor V.

A/Reference number: A42344; MUID:92232668; PMID:1567832

A/Accession: A56172

A/Molecule type: DNA

A/Residues: 1-2224 <CRT>

A/Cross-references: GB:J03368

A/Accession: A42344

A/Molecule type: DNA

A/Residues: 48-58;79-89;120-130;191-201;238-249;313-323;368-378;428-437;461-471;533-542;

R/Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm

Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987

A/Title: Complete cDNA and derived amino acid sequence of human factor V.

A/Reference number: A28028; MUID:87260886; PMID:3110773

A/Accession: A28028

A/Molecule type: mRNA

A/Residues: 1-857; 'R', 859-864, 'R', 866-924, 'E', 926-1763, 'I', 1765-2212, 'T', 2214-2224 <JEN>

A/Cross-references: GB:M18967

A/Note: Parts of this sequence, including the amino end of the mature protein, were dete

R/kane, W.H.; Ichinose, A.; Hagen, F.S.; Davies, E.W.

Biochemistry 26, 6508-6514, 1987

A/Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum

A/Reference number: A27498; MUID:88107860; PMID:2827732

A/Accession: A27498

A/Molecule type: mRNA
A/Residues: 1-1284, 'I', 1286-1600 <KAN>
A/Cross-references: GB:M17785
A/Note: parts of this sequence were determined by protein sequencing
R/Kane, W.H.; Davies, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A/Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homolog
A/Reference number: A25897; MUID:86313685; PMID:3092220
A/Accession: A25897
A/Molecule type: mRNA
A/Residues: 1188-1215, 1315-2224 <KA2>
A/Cross-references: GB:M14335
A/Note: parts of this sequence were determined by protein sequencing
R/Keller, F.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A/Title: Thrombin-catalyzed activation of recombinant human factor V.
A/Reference number: A56139; MUID:95210278; PMID:7696276
A/Contents: annotation; thrombin cleavage sites
C/Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C/Genetics:
A/Gene: GDB:F5
A/Cross-references: GDB:119396; OMIM:227400
A/Map position: 1Q23-1Q23
A/Intons: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 65/
C/Function:
A/Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot;
A/Pathway: blood coagulation
C/Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re;
F/1-28/Domain: signal sequence; duplication; glycoprotein; phospholipid binding; plasma;
F/29-2224/Product: coagulation factor V #status predicted <SIG>
F/29-737/Product: coagulation factor V #status predicted <MAT>
F/29-345/Domain: A1 <DA1>
F/33-329/Domain: ferroxidase repeat homology <FO1>
F/346-691/Domain: A2 <DA2>
F/692-1573/Domain: B <DOB>
F/1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F/1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F/1574-1905/Domain: A3 <DA3>
F/1667-1765/Region: ferroxidase repeat homology <FO3>
F/1906-2064/Domain: C1 <DC1>
F/2065-2224/Domain: C2 <DC2>
F/2065-2221/Domain: discoidin I amino-terminal homology <DN1>
F/2065-2224/Domain: C2 <DC2>
F/2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F/51-55,235,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,149/
F/167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/bisulfide bonds: #status
F/334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F/363-693,1546/Binding site: sulfate (Tyr) (covalent) #status predicted
F/376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F/382-1338/Binding site: carboxylate (Asn) (covalent) #status predicted
F/534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F/737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F/1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F/1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 22.8%; Score 2824.5; DB 1; Length 2224;
Best Local Similarity 30.5%; Pred. No. 3e-139;
Matches 768; Conservative 390; Mismatches 850; Indels 507; Gaps 77;
QY 22 RRYLGAVELSWDMQSDGLBLPVDARFPVPKSPFPNTSVV-YKTLFVEFDHLFNI 80
Db 32 RQFVAAQGISWRPE-----PINSLSLSTVTFKTIIVREDFY-FKK 75

QY 81 AKRPPPMGLLGTQIAQVYDVTWITLKNMASHPVSLHAVGVSYMKASEGAEYDDQTSOR 140
Db 76 EKQSTISGLLGLTYLAVGVGLIKVFKKADKPLSLHPQIRYKSKSEGASYLDHTFPA 135
QY 141 EKEDKVPFGGSHYVWQVLKENGMSADPLCLTYSYLSHVDLVKDLNSGLIGALLVCRE 200
Db 136 EMDDAVAPGREGTYEWSISDSGTHDDPCLTHIYVSHENLIEDFNSGLIGLICK 195

QY	320	MDLQQLSHCHSHQHDGMEAYVYVSCPEEPOLIMKNEEAEYDDDLTDSMDVVRP	379	Db	1195	QPSISPDGLQWALSPDPQESLSPDLGQ	TSLSPDLSQESLSPDLGQTA	1242
Db	301	SPEGSWIASLPRHFOAGMOAYIDIKNCAKTRPKK	343	QY	1334	LKQFRLPLEBT	STLTQIDVNEKKGAIQ	1386
QY	380	DDNSPSFIQIRSVAKHPKTHVHIAAEEDWDYAPLVLAPDDRSYKQSYLNNGPQIRG	439	Db	1243	LSPD	QESLSPDLGQALSPDPS	1294
Db	344	RRHKKEWFIABEVEWYAPILPANMDKYSRSLHLDNFSNRIG	388	QY	1387	SPISDCLTR	SHSIQANRSPULPIAKVSPSPSIRPIYILTRVLFDQNSSHLPAASYRKK	1443
QY	440	RKYKVRPMAYTDETFKR	497	Db	1295	TSLSPDLSQESLSPDLGQALSPD	QESLSPDLGQALSPD	1344
Db	369	KHYKVVYKQCDDEFTKELEDPSSGEGILCPITRAQVRDTLKIIVFNKMASRYSIYPH	448	QY	1444	DSGVQBSHSHLOKAKKNLSIALITLBTMTGQRBVGSGLTGSATNSVYKVKVENTVLKPPD	1503	
QY	498	GIT	548	Db	1345	DLGOTSISPD	GOESLSPDLGQALSPDPSQESL	1381
Db	449	GVTSPYDNEVNSSSTSGSNTMIRAVR	500	QY	1504	LPKTS	CKVILLPKVHIYQDLFPFTFNSGSPGHLIDVEGSLLOGTGEGAIKNEAN	1558
QY	549	TRYYSFVNMEEDLASLIGLIPLLICVBSVDQGNQIMSDKNVILFSPDENRSWYTE	608	Db	1382	LQOTSISPDLSQESLSP	DLGQALSPD	1418
Db	501	TEPYYSNVDTITRLASGLIGLILLIKSRLDRGIQRAADISQAVFAVDENKSWYIED	560	QY	1559	RPGKVPFLRVATESSAKTPSKLLDPLAWNHYGTQIPKEEWKSQESPEKTAFKKDTI	1617	
QY	609	NIQRELPNPAQVLEDPPEFQASNMH	662	Db	1419	GOPL	SPDLSLES	1445
Db	561	NIYKCENEKVKRODPKXVESNMSNFTLPAINGVVPESIPIIGFCEPDIVQMFCSVG	620	QY	1618	LSLN	ACESHAIJAINEGQNKPEIEVTWAKQGRTERLCQ	1665
QY	663	AOTDFLSVFFSGVTFKHMVYEDTLTLPFSGETVFMSENPGLWILGCHNSDPRNMGMT	722	Db	1447	PDLNQTSHTSSSSSLSPLPFGQT	FPNADIQOMSPSPDS	1506
Db	621	TQNDILTIHFTGSHSFIYKGRHEDTLTFPMQGESVTIVMDNVGTWMLITMNSNPRSKKL	680	QY	1666	QRE	ITRTTLOSQDEB	1707
QY	723	ALLKVSCKNTGVYEDSYEDI	755	Db	1507	SRDDGDYI	PROKESSEEDYGBFEFVAYNDPYQTLRT	1560
Db	681	LRPRDAKLRNDD	737	QY	1708	RSPOKTRHYFAAVERLMDYGMSSPHVLRNRAQSGS	VPO	1758
QY	756	IEPRFSQNSRHPSTKQONATIP	814	Db	1561	WYLRNSTGNRKYYITAAEISWDYS	KFVQSDVDYVPEDIVYKVKYRKYL	1611
Db	738	LGLRSFNSSL	783	QY	1759	DSFTQLYRGELNEHLLGILGPVIRAEVDNIMVTFRNOASRPYSYSSLSIYE	1812	
QY	815	QSTPHCLSLSDQEAKEVFSDPSGA	864	Db	1612	DSFTKLDPOGVEEHLGILGPVIRAEVDVQVRFKNLASRPSYSLHAHGLSYEKSEK	1671	
Db	784	RSHVSRILAKFAESLTLHLERAPAGSPLERAGLDKNSALNP	838	QY	1813	EDORQGAEP	RKNVFKNETKTYFWKVOHMAPTKDEBCKAWAYFSVDLEKDVHSLG	1870
QY	865	DMVFTPEGLOL	904	Db	1672	TYEDDSEPMFKEDNAIQNKTYTYVWEATTSRSPENPGSACRAWAYYSANPEKD	1731	
Db	839	DPREDHPLSDVTGVSLLPFGTFGRNRPKXQRFQVGRGQAACHKFSQTPPAKTRTL	898	QY	1871	IGPLLVCHTNLNPAGHQVTVQESFALFTIPEBTKSWYFTEMNERNCRAPCNIOMEDPT	1930	
QY	905	ISTIPSDNLAAHTNTGS	954	Db	1732	IGPLLCRKTLDEKNMPVDMMREFVLLFVFPDEKKSYYDKKPTRSWRASS	1785	
Db	899	SDQNSSSSRMCP	942	QY	1931	FKNYRPHALNGYIMDTPLQVLMAQDQIRWYLLISMGNSNIHSIHPSGHVTVVRKEEY	1990	
QY	955	LSLSENNDSKLESGLMNSQES	1004	Db	1786	VKSHEFHAINGLYN	LPGLRMYEQEWVRLLELNLGSRDIDHVVHPHGTLENGTQQH	1844
Db	943	YEIITQDANENKTVNK	997	QY	1991	KMALNLYPGVFTVEMLPDSKAGIWRVVECLIGELHAGMSTLFLVYNSKQCTPLGMASGH	2050	
QY	1005	KDNALFVKSISLLKTNKTSNNSATNRXTHIDGSPILLIENSVPVQ	1054	Db	1845	QLGWPLLPSPGFKTEMLKASKPGWLLDTEVEIQIRAGMQTPFLIVDRECKMPGLSTGL	1904	
Db	998	DRQERNRSLKEGLLR	1048	QY	2051	IRDFQITASGOYGOAWPKLAELHYSGSINAW	STK	2104
QY	1055	DTEFKKVTPLIHDRMLMDKNATALRNLHNSKNTTSSKNMVMQCKEPIEPDAQNPMS	1114	Db	1905	IADSIQOASBEPGWGWEPLARLNGGSGYNAMIAEKLSTERNPEPFIQVDMQKEVLLTGIQ	1954	
Db	1049	DRHN	1087	QY	2105	TOGAROKPSSYISQFIIMYSIDGKQWQTVGNSGTGLMVFNGVDSGICKHNFNPLI	2164	
QY	1115	FFMFLPESARWIORTHGKNSLN	1171	Db	1965	TOGAKHYLKPYTTEFCVAYSILDRKWRI	FKGNSRNVNVPFGNSDASTIKENQIDPPV	2024
Db	1088	HOQTSNDTTSQTSPPDLIVTVSPPEHY	1117	QY	2165	ARYRLHPTHYSIRASTLRMBELMGCDLASCMSPLGWSKALISDAQTASSYTNMFAT	WS	2223
QY	1172	GRGEFTKDYGLKEMVPPSSRNILFLNLDNLHNNHTNQEKKEQEBIEKKEETLIQENVLP	1231	Db	2025	ARYIRISPTGVSXPKPALALEQGCVEVNGCSTPLGNSCKIENKQITASSPKSKWGNWE	2084	
Db	1118	IFP	1139	QY	2224	PSKABLLHQSNSNARPOVNNPKELQVDFOKTMKVTGVTQGVKSLTSMYVKEFLSS	2283	
QY	1232	QIHVTGTGNPMKMLFLSTRQNEGSYD	1273	Db	2085	PLAELNAGRVNACAKANNNNQWLQDLKIKITAIIVTOGCKSLSESMYKSYTHY	2144	
Db	1140	PTHSTTAPSNRSP	1194	QY	2284	SQDGHQWTLFFQNGKV	KVFOGNODSFTPVVNSLDPPLLYRLHQPQSWHQIALRME	2341
QY	1274	NDSTNRKGIHAFSKGBEENLEGLNQTKQIVKEYACTTRISPNTSQCNFVTRSKRA	1333					

Db 2145 SDQTDWKP*REKSSMVDKIFEGNNVRGVHVFNFPIISRFIRIIPKTKWQSIARLE 2204

Qy 2342 VLGCETADLY 2351

Db 2205 LFQC---DMY 2211

RESULT 7

A25945

coagulation factor VIII - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000

C:Accession: A25945

R:Toole, J.J.; Pittman, D.D.; Orr, E.C.; Murtha, P.; Wasley, L.C.; Kaufman, R.J.

Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942, 1986

A:Title: A large region (approx 195 kDa) of human factor VIII is dispensable for in vitro

A:Reference number: A25945; MUID:86287369; PMID:3016730

A:Accession: A25945

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-869 <TOO>

C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase

Query Match 17.6%; Score 2190; D3 2; Length 869;

Best Local Similarity 45.2%; Pred. No. 1.1e-106;

Matches 493; Conservative 122; Mismatches 251; Indels 224; Gaps 15;

Qy 705 GLWILGCHNSDFRNRGMTALLKVSSCDKNTGYEDYSYDLSAYLLSKNALPERSFSQN 764

Db 1 GLWVLGCHNSDLNRGMTALLKVYSCDRDTGDDYDNTYEDIPGFLLSGKNVIEPRSPAQN 60

Qy 765 SRHPSTROQFNATTIPENDIEKTDWPAHRTMPKIQNVSSDDLMLLRQSTPHGLSL 824

Db 61 SRPESASQOFITTSPEDDVE-LDPQSQRTOALELSVPSGSGMLLGONPAPHGSS 119

Qy 825 SDIOEAKYETSPDGPATDSNLSSEMTFRPQLHSGDMVFTPSGILQLRLNEKLT 884

Db 120 SDIOEARNE--ADYLPGARERNVTAASAAARPELHHSERVLTEP-----EK--- 167

Qy 885 TAATEKGLDFKVSSTNNLIS--TIPSDNLAAGTNTSSLGPPSPMPVHVSQDITLFG 942

Db 168 ----ELKLDKSKSSSDLLKTSPTIPSDLSAETERTHSLGPPHPQVNFPSQIGAILVG 223

Qy 943 KKSPLTESGGLSLSENNDKSLGSLMNSQESSGKNSVTSGLPKGKRAAGPAL 1002

Db 224 KNSSHFTGAGVPLGSTEDH-----ESSLGENVSPVESGICEKRAHPAS 270

Qy 1003 LTKDNALFKVISILKTKNTKSNNSATNKTCHIDGPSLLIENSPPVWQNILESDFPKVT 1062

Db 271 LTKDDVLFKNISLVKTKKARVYLKTKRKHIDDAALLTENRASA----- 315

Qy 1063 PLIHDRMLKDNATALRLNENSKNTTSKKNMVMQKKEGIPPDPAQNDMSFFKMLPLP 1122

Db 316 -----TFMDKNTTASGLNHYVN----- 332

Qy 1123 ESARKIORTGKNSLNSQSGSPKQLVSLGPEKSEVGNFSLSEKNVVGKGEPTKDVGL 1182

Db 333 ---WIKGPLGNPLSBERGSEPELLTSSGSKSVKQSSQGRIRINAVEEELSKG--- 385

Qy 1183 KEMVPPSSRNPLFLNLDNLNHNTHNQEKKEIEKETTILQENVVLPIQHTVTGKNF 1242

Db 386 KEMMLPNSLTLFLNSADVQNDTHSQGKSEWREKLVQPKVDLPQVVTATGKNF 445

Qy 1243 MNKFLLSLSTRQNEGSYGAVPLVQDFRSLNDSNTNRKTKHTAHPKSGBEENLGLGNQ 1302

Db 446 LRNIHFQSTEPSVEGSDGSHAPVQDGRSLNDSAEAEATHAHSATREAPLEAPGNF 505

Qy 1303 TKQIVEKVACTRISPNTSQQNFVQTSKRAKQFRLPLETELEKRIIVDDTSQWSKN 1362

Db 506 T-----GPGPRSAVPRVKQSLKQIRLPLEEKIPRGVVLNATIRWS-- 548

Qy 1363 MKHLTPSTLTQIDYNEKEKGAITQSPSLDCLTRSHSIPOANRSPLPKIAKVSFPISREIY 1422

Db 549 ----- 548

Qy 1423 LTRVLPQDNSSHLPAASVRKKGSGVOESSHFLOGAKKNNLSLAILTLEMTGDQREVGSGL 1452

Db 549 -----ESSPLQGAKKNNLSLPFLTLEMAAGGQKISALG 582

Qy 1483 TSATNSVYKKVENTVLKPDLPKTSKGVLLPKVHYQKDLFFETTSNGSPGHLDLVEG 1542

Db 583 KSAAGPLASGKLEKAVLSSAGLSEASGKAEPLKVRVEREDLLPKTSNVSCHAGDLGQE 642

Qy 1543 SLLQGTGALKWNEANRPKVPFLRVATESAKTPSKLLDPLANDNHVGTQIPKEEWSQ 1602

Db 643 IFLOKTRGFVNLRKVRFG-----RTPSKLLGP-----PMPK-EWESL 679

Qy 1603 EKSPEKTAFAKKDII-LSLNACSNHAAINEGONKEPIEVTWAKQGRTERLCSQNPVP 1661

Db 680 EKSPASTALRTKDIISLPLDRHESNHSIAAKNEGOAETQREAAWTKQGPQRLGAPKPPV 739

Qy 1662 LRRHQREITRTTQSQDBEIDYDITTSIVEMKKEDPDIDYDENQSPRSFQKTRHYFAA 1721

Db 740 LRRHQREISLPTROPEEDKMDYDDIFSTETKGEDEFDIYGEDENQDPRSFQKTRHYFAA 799

Qy 1722 VERLDYGMSSSPHVLNRNRAQSGSVPOKQKVVQFETDGSFTQPLRYGELNEHGLGLQY 1781

Db 800 VEQLWDYGMSSSPALNRNRAQNGVEPRFKKVVFRBRADGSGFNPYRGELNKGHLGLQY 859

Qy 1782 IRAEVEDNIM 1791

Db 860 IRAEVEDNIM 869

RESULT 8

KUHU

ferroxidase (EC 1.16.3.1) precursor [validated] - human

N;Alternate names: ceruloplasmin

C;Species: Homo sapiens (man)

C:Date: 31-Aug-1980 #sequence_revision 12-May-1995 #text_change 08-Dec-2000

C:Accession: A25443; A35450; A00524; I59057

R:Koshchinsky, M.L.; Funk, W.D.; van Cost, B.A.; MacGillivray, R.T.A.

Proc. Natl. Acad. Sci. U.S.A. 83, 5086-5090, 1986

A:Title: Complete cDNA sequence of human preceruloplasmin.

A:Reference number: A25443; MUID:86259737; PMID:2873574

A:Accession: A25443

A:Molecule type: mRNA

A:Residues: 1-1060,1065-1069 <KOS>

A:Cross-references: GB:M13699; NID:G180255; PIDN:AAA51976.1; PID:G-80256

A:Note: this is the short or CP-2 alternatively spliced form

R:Mercer, J.F.B.; Grimes, A.

FEBS Lett. 203, 185-190, 1986

A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-terminal lead

A:Reference number: A24165; MUID:86275241; PMID:3755405

A:Accession: A24165

A:Molecule type: mRNA

A:Residues: 1-401549-599,784-829,919-952 <MBE>

R:Yang, F.; Friedrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Horton, W.A.

J. Biol. Chem. 265, 10780-10785, 1990

A:Title: Human ceruloplasmin. Tissue-specific expression of transcripts produced by alte-

A:Reference number: A35450; MUID:90285218; PMID:2355023

A:Accession: A35450

A:Molecule type: DNA

A:Residues: 1007-1064 <YAN>

A:Cross-references: GB:J05506

A:Note: this is the long or CP-1 alternatively spliced form

R:Takahashi, N.; Ortel, T.L.; Putnam, F.W.

Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984

A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid sequence

A:Reference number: A00524; MUID:84119493; PMID:6582496

A:Accession: A00524

A:Molecule type: protein

A:Residues: 20-1060,1065-1069 <YAK>

A:Note: 79-Gly and 49-Gly were also found

R:Yang, F.; Naylor, S.L.; Lum, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, K.H.; McGill

Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
A:Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
A:Reference number: I59067; MUID:86205876; PMID:3486416
A:Accession: I59067
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 218-1059
A:Cross-references: GB:M13536; MID:g180248; PIDN:AA51975.1; PID:g180249
A:Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions per molecule. In Wilson's disease the plasma levels of ferroxidase are diminished or undetectable. The three fragment chains are produced spontaneously during purification and are not artifacts.
C:Genetics:
A:Gene: GDB:C2
A:Cross-references: GDB:119069; OMIM:117700
A:Map position: 3q23-3q25
A:Introns: 106/3; 1061/1
A:Notes: the list of introns is incomplete
C:Function:
A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the reduction of NADPH to NADP+.
A:Note: other possible functions are amine oxidase activity, copper transport and homeostasis.
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidoreductase; signal sequence; status predicted <MATH>
F:1-19/Domain: signal sequence
F:20-1069/Product: ferroxidase, long form #status predicted <MATH>
F:20-1060/Product: ferroxidase, short form #status experimental <MATH>
F:20-499/Product: ferroxidase 67K chain #status experimental <K67>
F:23-357/Domain: ferroxidase repeat homology <F01>
F:373-718/Domain: ferroxidase repeat homology <F02>
F:501-905/Product: ferroxidase 50K chain #status experimental <K50>
F:733-1059/Domain: ferroxidase repeat homology <F03>
F:907-1065/Product: ferroxidase 19K chain #status experimental <K19>
F:138,397,762/Binding site: carboxylate (Asn) (covalent) #status experimental
F:174-200,276-357,534-560,637-718,874-900/Disulfide bonds: #status predicted
F:227,588,926/Binding site: carboxylate (Asn) (covalent) #status absent
F:235,338,343/Binding site: copper (His, Cys, His) (type 1) #status predicted
F:358/Binding site: carboxylate (Asn) (covalent) (partial) #status experimental
F:656,699,704,709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
F:994,1040,1045,1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 10.7%; Score 1326; DB 1; Length 1069;
Best Local Similarity 19.2%; Pred. No. 2.6e-61;
Matches 399; Conservative 204; Mismatches 419; Indels 1054; Gaps 23;
QY 5 LSTOFFCLLPFSATRRYVLGAVELSDWYQSDLGE--LPVDARPPRPVPSFPENT 61
DB 6 LGIFLFLCSTP-AWAKEKHVYIGIETWDY-ASDHGKKLISVDTEHSNTYLONGPDR 63
QY 62 SVVYKTLFVETDLENIAPRPPWMLGLFTQAEVYDVTWITLKNMASHPVSLHAGV 121
DB 64 GRLYKKALYQITDETFTTEKTPWLGFLGPIIKAEIGDGVYVHLKXLASRPYTFHSHG 123
QY 122 VSYWKASGAEYDDQTSOREKEDDKVFGGSHYVQVYKENGPMASDPLCLITYLSHV 181
DB 124 ITTYKEHEGAYPDNTDFQADRVYQEQYTYMLLATEQSPGEGGNCVTRIYHSHI 183
QY 182 DLVQNLNGLICALLVCEGSLAKSKTQL-HKFILLFAVDFECKSWHE-----TKNS 234
DB 184 DAPKDIASGLIPLICKDSDKREKHIDREFVWMSVDENFSWYEDNIKTYCSEP 243
QY 235 LMQDRDAASARAWPMKMTWGVYVNRSLPGLIGCHRSKVYVHVMGTTPVHISFLEHT 294
DB 244 EKVDKDNEDFQSNRMYSVNGTTFGLSLPGLSCAEDRVKWLFGMGNEVDVHAFFHQA 303
QY 295 FLVNRHQASLEISPIITFTLAQTLMDLQFLSCHISHQHDGMEAYVKVDSCEEPQL 354
DB 304 LINKVRIITNLFPATLDAYWAQNPGEWMLSCQNLNEKAGLQAPFQVQEC----- 357
QY 355 AKKNEEAEYDDDLTSDENVVRFDDNSPSFIQIRSVAKKPKTWVHIAEEEDWDY 414
DB 358 ---NKSSSD-----NIRKXVH-----YIIAEELIWN 385
QY 415 APL-----VLAPDRSYKQYLLNNGPQIRGRKYKVRFMAYTDET-----KTREA 462

386 APSGIDIFTKENLTPAGSDS--AVFFQGTTRIGSYKKLYREYTDASFNNRERGPBE 443
QY 463 HESGILGPLYGEGYDGLLIIFKNOASRPYNIYPHGI-----TDVRPLY---SRRLP 511
DB 444 EHLGILGPIVWAEVGDTRIVTFHNKGAYPLSIEIGVRFKNNEGTYYSNPYNQSRASVP 503
QY 512 KGVHKLDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSVFNMERDLASGLIGPLL 571
DB 504 PSASH-----VAPETFTYEWTPKEVGTPNADPVCLAKTYSAVDPTKDTFTGLIGPMK 558
QY 572 ICYKESVDCKNQMSDKRNVILSVFDPENRSWLTENIQBELPNPAGVQLEDDEFOASN 631
DB 559 ICKGSLHANGREQDVKDFEYLFPTVFDENESLLEDNIRMTTAPDQVDEKDEFOASN 618
QY 632 IMHSINGYVEDSLQ--LSVCLHEVAYWYILSICAQTDFLSVFSPGYTFKHKMYEITLIF 690
DB 619 KXHSWNGFWYGNQPGLTWCKGDSVWYLFSAAGNEADVHGIYFSGNTYLMRGRERRDTANLF 678
QY 691 PRSGETVFMENPGWLILGCHNSDFNRGMTALLKVSCKDXTGTYEDSDYEDSAYILL 750
DB 673 POTSLLHWPDETEGTENVECLTTHYTGKMKQKVTNQ-----718
QY 751 SKNAIEPRSFQNSRHSSTROKQFVATTIPENDIEKDPWFHRTPEMPKIQNVSSDILL 810
DB 719 ---PRQSE-----DSTFYLGERTYI-----735
QY 811 MLLRQSPTPHGLSLDLQEAQYETSDDPSPGALDSNLSLSEMTHTFRPOLHHSQGMVFTP 870
DB 736 ---IAAVEVE-----742
QY 931 HYDSQDLDTLFGKKSPLTESGGLSLEENNDKLESLMNSQESSWGKXVSTESGR 990
DB 743 -----742
QY 991 LFRKGRAHPALLTKDNALFKVSIKTKNTKNSATNRKTHIDGFSLLIENSVPWQN 1050
DB 743 -----WD-744
QY 1051 ILESDTEFKKVTFLIHDRMLMDKNAFALRNHMSKNTSSKNMEMVQKKEGIPPPDAQN 1110
DB 745 -----744
QY 1111 PDKSFFKWLFLPESARWIQTHGKNSLNSGGPSPKQVLSLGPESKVEGQNFLEKXKV 1170
DB 745 -----744
QY 1171 VGKGETRQVGLKEMVFPSSRNLFITNLNLHENNTHNQEKIOEBIEKKEKTLIQENVVL 1230
DB 745 -----744
QY 1231 PQIHTVTGKTNFKNKLFLSTRQNVGSDGAYAPVLDQFSLNDSTNRTKKHTAHSFKK 1290
DB 745 -----YSP-----747
QY 1291 GEEENLEGLNQTKQIVEKYACTTRISDNTSQNFVQSRKALKQFELPLETELEKRI 1350
DB 748 -----QR-----749
QY 1351 IVDDTSTQWNRKHLTPSTLTQIDYNEKEGATQSPSLDCLTRSHSIPQNSRPLPIA 1410
DB 750 -----EWEKELHHL-----758
QY 1411 KVSPPSIRPIYLTFLVFDQNSSHLPAASRYAKKDSGVQESSHFLQAKKNLSLAILTLE 1470
DB 759 -----QEQVSNRAFL-----768
QY 1471 MTGQREVGLSGTSSATNSVYKKVNTVLPKPTSGKVELLPKVHIYQKDLFPPTTS 1530

Arch. Biochem. Biophys. 293, 1-8, 1992
A;Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human
A;Reference number: S21692; MUID:92117681; PMID:1531003
A;Accession: S21692
A;Molecule type: protein
A;Residues: 20-29; Q: 902-910 <RYA>
C;Superfamily: ferroxidase; ferroxidase repeat homology
C;Keywords: copper; glycoprotein; oxidoreductase; plasma
P;1-19/Domain: signal sequence #status predicted <SIG>
P;20-1059/Product: ferroxidase #status predicted <MAT>
P;23-356/Domain: ferroxidase repeat homology <FO1>
P;372-712/Domain: ferroxidase repeat homology <FO2>
P;727-1053/Domain: ferroxidase repeat homology <FO3>

Query Match 10.3%; Score 1277; DB 1; Length 1059;
Best Local Similarity 19.1%; Pred. No. 9.3e-59;
Matches 395; Conservative 196; Mismatches 424; Indels 1054; Gaps 29;

Qy 5 LSTCFPLCLRFPCSTRYILGAVEJLWDMQ-SDLGEL-PVDARFPVRPKSEFNTS 62
Db LSALLFL-HSSLAWTREKHVYIGITEAVWDYASSEEKELISVDTOSNFVLRNCPDRIG 64
Qy 63 VVYKKTLPVEFTDHLNIAKRPBPMGLGPTIOAEVDTVITLKNMASHPVLSHAGV 122
Db 65 RYKXKALYSSTYDGTFTKIDKPAWGLFGLPVIAEVCVKVSHVKNFASRYTFHAGV 124
Qy 123 SYMKASGAZYDDQTSQREKEDKVPFGGSHYVQVVKENGPMASDPLCLTYLSHVD 182
Db 125 TYTKANEGALYDNTTDFQRADKLFGQQLYVLR-NEPSPGEGDSNCTRIYHSHVD 183
Qy 183 LVKOLNSGLIGALLVCREGSLAKETQL-HKFTLLFAVDFEGKSHSETKNSLM----- 236
Db 184 APKDIASGLIGLILCKGSLHKEENIDQEFVLMFSVVDENLSWLEDNIKTFCSBPE 243
Qy 237 -QDRDAASAWPKVHTVGNVNRSLPGLIGCHRSVYVHWIGMTTPEVHSIFEGHTF 295
Db 244 KVDKDNEDFQESNEMYSINGYTFGLSLGSMCAEDRVAKYFLGMEVDVHSAFHGOAL 303
Qy 296 LVRNHRQASLEISPTITFLTAQFLMLMDQLFSLCHSHQHDGMEAYKVQSCPEEPQLT 355
Db 304 TSKNTHYTIINLFPATLIDVSMVAQNPGVWMLSCQNLNHLKAGLQAFQVRDC----- 356
Qy 356 MKNNEAEDYDDLDITSEMDVVRFDNDSFSLQIRSAKHPKTHVHYIAAEEEDWDYA 415
Db 357 --NKPSP--DDIQRHV-----RH-----YIIAEEITWDYA 385
Qy 416 P-----LVLAPDRSYKSYQLNNGPQIRGKRYKVRFMAYTDTFP---KTRSAI 461
Db 386 PGGTDTFTGENTSLGSDSRVFEQ---GATKIGGSKYKLVREYTDSDFTNRKERGPD 441
Qy 462 QHESGILGPLLYGEVGDITLLIFQKQASRPYNIYPHGITDVR-----PLYSRLPKGVKHL 517
Db 442 EEHLGILGPVIAEVDGDIIRVTFHNGQPLGPIQPMGVFTKENEGTYG---PDGRSSK 498
Qy 518 KDFPILPGRIKPKVITVYEDGPTKDPCLTRYSSFVNMRDLASGLIGLILCYKES 577
Db 499 QASHVAPKETFTYVETVPEMGTYADPVCLSKMYISGVDLTKDITFTGLIGPKICKGS 558
Qy 578 VDQKQGNISDKRNILFVSFVFNENRNYLTENFQFLPNPAGVQLEDPPFQASNIHSHIN 637
Db 559 LLADGRQKVDKEFFYLFAVTFDENESLLDDNTRMTTAPENVDVDEDEDQESNKHSMN 618
Qy 638 GYVDSIQ-LSVCLHEVAYVYIISGAQDTFLSVFSGYTFKHKVYVEDTLTLFPFSGET 696
Db 619 GFMYGNLPGNLMLGIESIVWYLFSGAGNEADHGVYFSGNTYLSKGERDRTANLFPKHSJT 678
Qy 697 VFMSENPGI-WILGCHNSDFRNGMTALLKVSSCDKNTGVDYSDYSDIISAYLLSKNNAI 756
Db 679 LLMTPTDEGDFVBECLTTHYTGKMKQYTVNQC-----KQGFEDVT----- 720
Qy 757 EPRFSQNSRSPETROKFNATTIPENDIEKTDPEFAHRTMPKIQNVSSDILLMLLRQS 816
Db 721 ----- 720

RESULT 9

A35210
ferroxidase (EC 1.16.3.1) precursor - rat
N/Alternate names: ceruloplasmin
C/Species: Rattus norvegicus (Norway rat)
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
C/Accession: A35210; A41753; A29564; S21692
R/Fleming, R.E.; Gitlin, J.D.
J. Biol. Chem. 265, 7701-7707, 1990
A;Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene exp
A;Reference number: A35210; MUID:90237081; PMID:2332446
A;Accession: A35210
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1059 <FL2>
A;Cross-references: GB:J05424
R/Fleming, R.E.; Gitlin, J.D.
J. Biol. Chem. 267, 479-486, 1992
A;Title: Structural and functional analysis of the 5'-flanking region of the rat cerulop
A;Reference number: A41753; MUID:92112697; PMID:1730611
A;Accession: A41753
A;Molecule type: DNA
A;Residues: 1-48 <FL2>
A;Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 a
R/Aldred, A.R.; Grimes, A.; Schreiber, G.; Mercer, J.F.B.
J. Biol. Chem. 262, 2875-2878, 1987
A;Title: Rat ceruloplasmin. Molecular cloning and gene expression in liver, choroid plex
A;Reference number: A29564; MUID:87137545; PMID:3818625
A;Accession: A29564
A;Molecule type: mRNA
A;Residues: 'NSG', 215-216, 'Y', 218, 'FAT', 222, 'F', 224-226, 'E', 228, 'LL', 231, 'D', 233-235, 'RM
A;Experimental source: liver
A;Note: the authors translated the codon GCG for residue 60 as Gly and GTG for residue 1
R/Ryan, T.P.; Grover, T.A.; Aust, S.D.

C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol
F;1-21/Domain: signal sequence #status predicted <SIG>
F;28-60/Domain: EGF homology <EG1>
F;68-107/Domain: EGF homology <EG2>
F;110-267/Domain: discoidin I amino-terminal homology <DN1>
F;271-427/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.3%; Score 663; DB 2; Length 427;
Best Local Similarity 40.3%; Pred. No. 2.9e-27;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;

Qy 2018 ECLIGELHAGMSTFLVYSNK-----COTPLGMASGHIRDFQITASQY-- 2062
Db 78 KCLVTEDTQRG--DIFTEYICQCPVGSIGHCEIGCCSTKLGEGGAADSOISASSVVMG 135

Qy 2063 ----GOWAPKLARLHYSGSINAW--STKEPSWIKVDLLAPMIHGIKTQAGKFFSLY 2116
Db 136 FMGLQRWGPELARLYRTGIVNANTASSYDVKPWIQVDFLRKRVSGVMVTQASAGRAEY 195

Qy 2117 ISQFLIMVSLDGKQWYRGNSTGLMVFVGVNDSGGIKHNIENPPIIARYIRLHPHYS 2176
Db 196 LKTEKVASVLDGRPEFTQDESGETGDKFMGQNDNNSLKINNFNPTLEAQYIRLLFVSCH 255

Qy 2177 IRSTLRMELMGCDLNSCMPLGMESKAISDAQITASSYFT--NMFA-TWSPSKARLHLQG 2233
Db 256 RGCTLRFELLGCELHGCEBPLGKNTIPDSQITASSYKTNLRAFQWYPHLGRDLNQG 315

Qy 2234 RSNAPRPQVNPKEWLOVDFOKTMKVGTGTVGVSKSLTSTMTVKEFLSSSQDGHQWTLF 2293
Db 316 KINAWTAQNSAKAWLQVDLQKQKVTGIIQTGARDFGHIQYVASYKVAHSDGQWTVY 375

Qy 2294 FQNGKVFQGNQDSFTPVNSLDPLRLYRIHPQSVWVHOIALRMEVLGC 2345
Db 376 EQQGTSKVFGQNLNDSNKKNIPEKPFMARYVRVLPLSWENRITLRVELLGC 427

RESULT 12
A36479
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36479
R;Stubbs, J.D.; Lektis, C.; Singer, K.L.; Bui, A.; Yuzuki, D.; Sriniwasan, U.; Parry, G
Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990
A;Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the ex
A;Reference number: A36479; MUID:91046006; PMID:2122462
A;Accession: A36479
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-463 <STU>
A;Cross-references: GB:M38337; NID:G199142; PIDN:AAA39534.1; PID:G199143
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homol
C;Keywords: membrane protein
F;68-107/Domain: EGF homology <EG1>
F;28-60/Domain: EGF homology <EG2>
F;147-303/Domain: discoidin I amino-terminal homology <DN1>
F;307-463/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.3%; Score 657; DB 1; Length 463;
Best Local Similarity 42.2%; Pred. No. 6.9e-27;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5;

Qy 2037 SNKCTPLGMASGHIRDFQITASQY-----GOWAPKLARLHYSGSINAW--STK2PFS 2088
Db 145 ASRCSTQLGMEGGAIAIDSQISASVYVMGFMGLQWGPPELARLYRTGIVNAWHAGSNYSKP 204

Qy 2089 WIKYDLAPMIHGIKTQAGKQKTSYISOFIIMVSLDGKQWYRGNSTGLTMVFPFN 2148
Db 205 WIQVNLKRVKRVSGVMWTCASRAGAEFLKTFKVAISLDGKFFIQDESQGD-KEFLGN 263

Qy 2149 VDSGKIHNIENPPIIARYIRLHPHYSIRSTLRMELMGCDLNSCMPLGMESKAISDAQ 2208
Db 264 LDNNSLKVMNENPTLEAQYIRLVPVSVCHRGCTLRFELLGCELHGCLPEPLGKNTIPDSQ 323

Qy 2209 ITASSYFT--NMFA-TWSPSKARLHLOGRSNAPVQVNPKEWLOVDFOKTMKVGTGTTQ 2265
Db 324 MSASSSYKTNLRAFQWYPHLGRDLNQGKINAWTAQNSAKEWLQVDLGFQRCQVTHITQ 383

Qy 2266 GVKSLTSTMTVKEFLSSSQDGHQWTLFFQNGKVKVQFGNQDSFTPVVNSLDPLLTRYL 2325
Db 384 GARDFGHIQYVESYKVAHSDGQWTVVEEGSGSKVFGQNLNDSNKKNIPEKPFMARYV 443

Qy 2326 RIHPQSVWVHOIALRMEVLGC 2345
Db 444 RVLFPVSWENRITLRVELLGC 463

RESULT 13
T11743
SP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 08-Sep-2002
C;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe:
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zo:
A;Reference number: 217325; MUID:98206817; PMID:9546740
A;Accession: T11743
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:G2652927; PIDN:CAA72379.1; PID:G2652928
A;Experimental source: testis
C;Function:
A;Description: may be involved in membrane remodeling and/or function as a zona pellucida;
C;Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo;
F;6-40/Domain: EGF homology <EGF>

Query Match 5.2%; Score 650; DB 2; Length 409;
Best Local Similarity 39.3%; Pred. No. 1.3e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

Qy 2018 ECLIGELHAGMSTFLVYSNK-----COTPLGMASGHIRDFQITASQY-- 2062
Db 58 ECEVIDDAHGG--DYVTEYICKPHGYTGHCETICNAPLGMETGALADPQISASSMHLG 115

Qy 2063 ----GOWAPKLARLHYSGSINAW--STKEPFWIKVDLLAPMIHGIKTQAGKFFSLY 2116
Db 116 FMGLQWAPELARLHRAHGINAWTASNYDRNPWQVNLRRMRVTGVTQGRASRAGSAY 175

Qy 2117 ISQFLIMVSLDGKQWYRGNSTGLMVFVGVNDSGGIKHNIENPPIIARYIRLHPHYS 2176
Db 176 MKTFKVAYSTDKRKPFQIGAEESGDKIFMGNLDSGLKVNLFVPLEVQYVRLVPIICH 235

Qy 2177 IRSTLRMELMGCDLNSCMPLGMESKAISDAQITASSYFTN---MFATWSPSKARLHLQG 2233
Db 236 RGCTLRFELLGCELSCABPLGLKNTIPNKQITASSFYRTNGLSFAFSWYFFYARLDNQG 295

Qy 2234 RSNAPRPQVNPKEWLOVDFOKTMKVGTGTVGVSKSLTSTMTVKEFLSSSQDGHQWTLF 2293
Db 296 KFNAWTAQNSASENIQIDLGSRQRRVTGIIQTGARDFGHIQYVAAKVAISDDGVSWTEY 355

Qy 2294 FQNGKVFQGNQDSFTPVNSLDPLLTRYIRLHPQSVWVHOIALRMEVLGC 2345
Db 356 RDQALGKLFPGNLNDSNKKNIPEKPFTRVRLPVAWHNRITLRVELLGC 409

RESULT 14
S65138
Glycoprotein antigen MG57/53, mammary gland - bovine (fragment)
N;Alternate names: Glycoprotein component 16/major fat-globule membrane protein/MFG-E8 hc
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C;Accession: S65138; G48394
R;Aoki, N.; Kishi, Y.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995

A>Title: Molecular cloning of glycoprotein antigens MGPS7/53 recognized by monoclonal anti
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-401 <NA>
R:Mather, J.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A>Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
II-like sequences.
A:Reference number: A48394; MUID:93250576; PMID:8485470
A:Accession: G48394
A>Status: preliminary
A:Molecule type: protein
A:Experimental source: milk
A:Note: sequence extracted from NCBI backbone (NCBIP:131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: glycoprotein
F1-32/Domain: EGF homology (fragment) <EG1>
F40-79/Domain: EGF homology <EG2>
F82-239/Domain: discoidin I amino-terminal homology <DN1>
F243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.1%; Score 635; DB 2; Length 401;
Best Local Similarity 37.9%; Pred. No. 7.8e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGHLHAGSTLFLVYSNK-----CQPLGASGHIRPQTASQY-- 2062
DB 50 ECQVTDSDHRG--DVFIQICPLGVGHIHCTTCTSPGLMOTGAIADSOISASSMHLG 107
QY 2063 ----GOWAPKLARLHYSGSINAWST---KEPFSWIKVDLLAPMIHGIKTQCARQKFS 2114
DB 108 FMGLQWAPLAEALHGTGIVNAWTSNGYDKNP--WQVNLMRKMWVTGVVTCASRAGSA 165
QY 2115 LYISQFIIMYSLDGKKWQYRGNSGTGLMVFQGVNDSSGKHNIPNPPIIARVIRLHPH 2174
DB 166 EYLKTFKVAYSTDGRQFQFQVAGRSQDKIFIGNVNSGLKINLFDTELETQVRLVPFI 225
QY 2175 YSTRILRMELMGCDLNSCMPLGMSKASDAQITASSYFTN--MFATWSFKARLHL 2231
DB 226 CHRGTCTRLPELGCNLGCTEPLGKNDTIPNKQITASSYKWTGLSASFSPFYARLDN 285
QY 2232 QGRSNARPOVNNPKMLQVDFQKTMKVTVGTTQGVKSLTSMYKVEFLISSQDGHQWT 2291
DB 286 QGFNFWAQTNSASEWLQIDLSQKRVGTIIQTGARDFGHIQVVAAYRVAAGDDGVWT 345
QY 2292 LFFQNG--KVKVFQGNQDSTFPVNSLDPELLTRYLRIHPQSWHQTALRMEVLGC 2345
DB 346 EYKDPGASESKIIFPGNNDNNSHKKNIFETPFQARFVRIQVAVHNRIILRVLLG 401

RESULT 15
PAS-6/7 protein precursor - bovine
N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E8 h
C:Species: Bos primigenius taurus (cattle)
C>Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text_change 04-Nov-2002
C:Accession: S74211; S78114; S24181; S65138; G48394
R:Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A>Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob
A:Reference number: S74211; MUID:97008954; PMID:8856064
A:Accession: S74211
A:Molecule type: mRNA
A:Residues: 1-427 <HVA>
A:Cross-references: EMBL:X91895; NID:G1632778; PIDN:CAA62997.1; PID:G1632779
A:Accession: S78114
A:Molecule type: protein
A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
R:Xim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992

A>Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
A:Reference number: S23926; MUID:92353107; PMID:1643094
A:Accession: S24181
A:Molecule type: protein
A:Residues: 383-394 <KIM>
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A>Title: Molecular cloning of glycoprotein antigens MGPS7/53 recognized by monoclonal anti
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-427 <AOK>
R:Mather, J.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A>Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-pig
II-like sequences.
A:Reference number: A48394; MUID:93250576; PMID:8485470
A:Accession: G48394
A>Status: preliminary
A:Molecule type: protein
A:Experimental source: milk
A:Note: sequence extracted from NCBI backbone (NCBIP:131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F1-18/Domain: signal sequence #status predicted <SIG>
F19-427/Product: PAS-6/7 protein #status experimental <MAT>
F24-58/Domain: EGF homology <EG1>
F56-105/Domain: EGF homology <EG2>
F108-265/Domain: discoidin I amino-terminal homology <DN1>
F269-427/Domain: discoidin I amino-terminal homology <DN2>
F24-35,29-47,49-58,66-77,71-94,96-105/Disulfide bonds: #status predicted
F27/Binding site: carboxylate (Ser) (covalent) #status experimental
F34/Binding site: carboxylate (Thr) (covalent) #status experimental
F59,227/Binding site: carboxylate (Asn) (covalent) #status experimental
F109-265,252-256,270-427/Disulfide bonds: #status experimental

Query Match 5.1%; Score 635; DB 2; Length 427;
Best Local Similarity 37.9%; Pred. No. 8.6e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGHLHAGSTLFLVYSNK-----CQPLGASGHIRPQTASQY-- 2062
DB 76 ECQVTDSDHRG--DVFIQICPLGVGHIHCTTCTSPGLMOTGAIADSOISASSMHLG 133
QY 2063 ----GOWAPKLARLHYSGSINAWST---KEPFSWIKVDLLAPMIHGIKTQCARQKFS 2114
DB 134 FMGLQWAPLAEALHGTGIVNAWTSNGYDKNP--WQVNLMRKMWVTGVVTCASRAGSA 191
QY 2115 LYISQFIIMYSLDGKKWQYRGNSGTGLMVFQGVNDSSGKHNIPNPPIIARVIRLHPH 2174
DB 192 EYLKTFKVAYSTDGRQFQFQVAGRSQDKIFIGNVNSGLKINLFDTELETQVRLVPFI 251
QY 2175 YSTRILRMELMGCDLNSCMPLGMSKASDAQITASSYFTN--MFATWSFKARLHL 2231
DB 252 CHRGTCTRLPELGCNLGCTEPLGKNDTIPNKQITASSYKWTGLSASFSPFYARLDN 311
QY 2232 QGRSNARPOVNNPKMLQVDFQKTMKVTVGTTQGVKSLTSMYKVEFLISSQDGHQWT 2291
DB 312 QGFNFWAQTNSASEWLQIDLSQKRVGTIIQTGARDFGHIQVVAAYRVAAGDDGVWT 371
QY 2292 LFFQNG--KVKVFQGNQDSTFPVNSLDPELLTRYLRIHPQSWHQTALRMEVLGC 2345
DB 372 EYKDPGASESKIIFPGNNDNNSHKKNIFETPFQARFVRIQVAVHNRIILRVLLG 427

Search completed: April 13, 2004, 14:09:55
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 18 seconds
(without alignments)
6800.941 Million cell updates/sec

Title: NP0000123-328-355-581

Perfect score: 12415

Sequence: 1 YQIELSTCFLLCLLPCEFA.....WVHQIALRMEVLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12399	99.9	2351	1 FAS_HUMAN	P00451 homo sapien
2	8820	71.0	2319	1 FAS_MOUSE	Q06194 mus musculus
3	8057	64.9	2133	1 FAS_PIG	P12263 sus scrofa
4	2823	5	2224	1 FAS_HUMAN	P12259 homo sapien
5	2776	5	2258	1 FAS_PIG	Q9G1P1 bos scrofa
6	2743	22.1	2211	1 FAS_BOVIN	Q28107 sus taurus
7	1322	10.6	1065	1 CERU_HUMAN	P00450 homo sapien
8	1274	10.3	1059	1 CERU_MOUSE	P13635 rattus norv
9	1243	10.0	1062	1 CERU_MOUSE	Q61147 mus musculus
10	663	5.3	427	1 MFGM_MOUSE	P70490 rattus norv
11	657	5.3	463	1 MFGM_MOUSE	P21956 mus musculus
12	650	5.2	409	1 MFGM_PIG	P79385 sus scrofa
13	635	5.1	427	1 MFGM_BOVIN	Q95114 bos taurus
14	588	4.7	387	1 MFGM_HUMAN	Q08431 homo sapien
15	469	3.8	931	1 NRP2_HUMAN	O60462 homo sapien
16	464	3.7	925	1 NRP2_MOUSE	O35276 rattus norv
17	462	3.7	931	1 NRP2_MOUSE	O35375 mus musculus
18	458	3.7	924	1 NRP1_MOUSE	P79795 gallus gall
19	451	3.6	922	1 NRP1_MOUSE	Q9G1J9 rattus norv
20	446	3.6	923	1 NRP1_MOUSE	P77333 mus musculus
21	443	3.6	928	1 NRP1_XENLA	P28824 xenopus lae
22	429	3.5	933	1 NRP1_HUMAN	O14786 homo sapien
23	306	2.5	3133	1 HMCT_BOVMO	P98092 bombyx mori
24	266	2.1	764	1 CPX2_MOUSE	Q9d215 mus musculus
25	261	2.1	756	1 CPX2_HUMAN	Q8n436 homo sapien
26	260	2.1	280	1 XLR1_FUGRU	P9W6R5 fugu rubrip
27	247	2.0	224	1 XLR1_MOUSE	Q92114 mus musculus
28	243	2.0	224	1 XLR1_HUMAN	O15537 homo sapien
29	242	2.0	3418	1 SRC2_HUMAN	P51587 homo sapien
30	235	1.9	3358	1 SIR4_YEAST	P11978 saccharomyc
31	223	1.8	722	1 CPXM_MOUSE	Q92100 mus musculus
32	221	1.8	3329	1 BRC2_MOUSE	P97929 mus musculus
33	219	1.8	1928	1 MYS1_YEAST	P08964 saccharomyc

RESULT 1
FAS_HUMAN
ID FAS_HUMAN STANDARD; PRT; 2351 AA.
AC P00451;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component)
DE (Antihemophilic factor) (AHF).
GN F8 OR F8C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86081164; PubMed=3935400;
RA Trnett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
Hartog K., Kuo C.H., Mesiarz F.R., Merryweather J.P., Najarian R.,
Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalaro J., Hansen J.,
Nordfang O., Ezban M.;
RA "Characterization of the polypeptide composition of human factor
VIII:C and the nucleotide sequence and expression of the human kidney
cDNA".
RT DNA 4:333-349 (1985).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061548; PubMed=6438526;
RA Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,
Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,
Delwart E., Ruddenham E.G.D., Vehar G.A., Lawn R.M.;
RA "Expression of active human factor VIII from recombinant DNA clones".
RT Nature 312:330-337 (1984).
RL [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85061550; PubMed=6438528;
RA Tootle J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Pass D.N.,
Hewick R.M.;
RA "Molecular cloning of a cDNA encoding human antihemophilic factor".
RT Nature 312:342-347 (1984).
RL [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=93265012; PubMed=1303178;
RA Gitschier J., Wood W.I.;
RA "Sequence of the exon-containing regions of the human factor VIII
gene".
RT Hum. Mol. Genet. 1:199-200 (1992).
RL [5]
RP SEQUENCE OF 2064-2070 FROM N.A.
RA de Water N.S., Williams R., Browett P.J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SULFATION OF TYR-1699.

P47580 mycoplasma
Q96sm3 homo sapien
Q9uhc6 homo sapien
P34216 saccharomyc
Q01484 homo sapien
Q01613 xenopus lae
Q10411 schizosacch
Q94887 drosophila
Q00799 plasmodium
P47549 mycoplasma
Q99p47 mus musculu
P20929 homo sapien

ALIGNMENTS

RX MEDLINE=91093266; PubMed=1898735;
RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
RA Xetters K., van Mourik J.A.;
RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RT essential for the interaction of factor VIII with von Willebrand
RT factor.";
RL J. Biol. Chem. 266:740-746(1991).
RN [17]
RP SULFATION.
RX MEDLINE=92207952; PubMed=154716;
RA Pittman D.D., Wang J.H., Kaufman R.J.;
RA "Identification and functional importance of tyrosine sulfate
RT residues within recombinant factor VIII.";
RL Biochemistry 31:3315-3325(1992).
RN [18]
RP STRUCTURE BY NMR OF 2322-2343.
RX MEDLINE=95200924; PubMed=7893714;
RA Gilbert G.E., Baleja J.D.;
RA "Membrane-binding peptide from the C2 domain of factor VIII forms an
RT amphipathic structure as determined by NMR spectroscopy.";
RL Biochemistry 34:3022-3031(1995).
RN [19]
RP REVIEW ON MOLECULAR BASIS OF HEMA.
RX MEDLINE=91221499; PubMed=1902642;
RA Gitschier J.;
RA "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
RN [10]
RP REVIEW ON MOLECULAR BASIS OF HEMA.
RX MEDLINE=89088506; PubMed=2491949;
RA White G.C. II, Shoemaker C.B.;
RA "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
RN [11]
RP REVIEW ON MOLECULAR BASIS OF HEMA.
RX MEDLINE=95245332; PubMed=7728145;
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RN Hum. Mutat. 5:1-22(1995).
RX [12]
RP VARIANT HEMA GLN-2326.
RX MEDLINE=86235434; PubMed=3012775;
RA Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
RA "Identification of a missense mutation in the factor VIII gene of a
RT mild hemophilic.";
RL Science 232:1415-1416(1986).
RN [13]
RP VARIANT HEMA PRO-2135.
RX MEDLINE=88056539; PubMed=3122181;
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
RA "A novel missense mutation in the factor VIII gene identified by
RT analysis of amplified hemophilia DNA sequences.";
RL Nucleic Acids Res. 15:9797-9805(1987).
RN [14]
RP VARIANT HEMA GLN-2228.
RX MEDLINE=88191889; PubMed=2833855;
RA Yousoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
RA Kazazian H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the
RT relative mutation rate at CG dinucleotides.";
RL Am. J. Hum. Genet. 42:718-725(1988).
RN [15]
RP VARIANT HEMA GLY-291.
RX MEDLINE=89220354; PubMed=2835904;
RA Yousoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
RA Antonarakis S.E.;
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
RT in exon 7 of the factor VIII gene.";
RL Am. J. Hum. Genet. 42:867-871(1988).
RN [16]
RP VARIANT HEMA CYS-1708.
RX MEDLINE=89274393; PubMed=2499363;
RA C'Brien D.P., Tuddenham E.G.;
RT "Purification and characterization of factor VIII 1,689-Cys: a
RT nonfunctional cofactor occurring in a patient with severe hemophilia
RL A.";
RN Blood 73:2117-2122(1989).
RX [17]
RP VARIANT HEMA CYS-391.
RX MEDLINE=90001543; PubMed=2506948;
RA Shima M., Ware J., Yoshioka A., Fukui H., Pulcher C.A.;
RA "An arginine to cysteine amino acid substitution at a critical
RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL Blood 74:1612-1617(1989).
RN [18]
RP VARIANT HEMA LEU-189.
RX MEDLINE=90057680; PubMed=2510835;
RA Chan V., Chan T.K., Tong T.M., Todd D.;
RA "A novel missense mutation in exon 4 of the factor VIII:C gene
RT resulting in moderately severe hemophilia A.";
RL Blood 74:2688-2691(1989).
RN [19]
RP VARIANT HEMA LEU-2326.
RX MEDLINE=89197216; PubMed=2495245;
RA Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
RA "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
RT of the factor VIII gene.";
RL Hum. Genet. 81:335-338(1989).
RN [20]
RP VARIANT HEMA HIS-391.
RX MEDLINE=89264602; PubMed=2498892;
RA Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Fujimaki M., Hoyer L.W.;
RT "Direct characterization of factor VIII in plasma: detection of a
RT mutation altering a thrombin cleavage site
RT (arginine-372-->histidine).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
RN [21]
RP VARIANT HEMA CYS-1708.
RX MEDLINE=90105723; PubMed=2104766;
RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
RT Cys) in the factor VIII gene of two unrelated patients with
RT cross-reacting material-positive hemophilia A.";
RL Blood 75:384-389(1990).
RN [22]
RP VARIANTS HEMA GLN-2228 AND LEU-2326.
RX MEDLINE=90123183; PubMed=2105106;
RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
RA Mancuso G., Votini M., de Biasi R., Baudo F., Carbonara A.;
RT "Recurrent mutations and three novel rearrangements in the factor
RT VIII gene of hemophilia A patients of Italian descent.";
RL Blood 75:662-670(1990).
RN [23]
RP VARIANT HEMA CYS-391.
RX MEDLINE=90329422; PubMed=1973901;
RA Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
RT internal heavy chain thrombin cleavage site.";
RL Br. J. Haematol. 75:73-77(1990).
RN [24]
RP VARIANTS HEMA PHE-1699 AND CYS-1708.
RX MEDLINE=90152691; PubMed=2105906;
RA Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,
RA Kazazian H.H., Antonarakis S.E.;
RT "Characterization of mutations in the factor VIII gene by direct
RT sequencing of amplified genomic DNA.";
RL Genomics 6:65-71(1990).
RN [25]
RP VARIANTS HEMA CYS-1728 AND ASP-1941.
RX MEDLINE=90169985; PubMed=2106480;
RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
RA Kazazian H.H.;
RT "Use of denaturing gradient gel electrophoresis to detect point
RT mutations in the factor VIII gene.";

Query Match									
Best Local Similarity 99.9%; Score 12399; DB 1; Length 2351;									
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;									
QY	1	MQIELSTCFCLLRFCATRYVYLGAVELSDYMQSDLGELPVDARPPVPKSPFN	60						
DB	1	MQIELSTCFCLLRFCATRYVYLGAVELSDYMQSDLGELPVDARPPVPKSPFN	60						
QY	61	TSVVYKTLFVEVTHLFIKAPPRPMMGLGPTIQAEVYDVTVIITLKNMASHPVSLHAV	120						
DB	61	TSVVYKTLFVEVTHLFIKAPPRPMMGLGPTIQAEVYDVTVIITLKNMASHPVSLHAV	120						
QY	121	GVSYKASGAEYDDQTSOREKDDKVFPQGSSTYVQVLKENGPKASDPLCLTYVLSH	180						
DB	121	GVSYKASGAEYDDQTSOREKDDKVFPQGSSTYVQVLKENGPKASDPLCLTYVLSH	180						
QY	181	VDLVKDLNSGLI GALLVCREGSLAKEKTOTLHKFILLFAVEDGKSWHSETKNSLMQDRD	240						
DB	181	VDLVKDLNSGLI GALLVCREGSLAKEKTOTLHKFILLFAVEDGKSWHSETKNSLMQDRD	240						
QY	241	AASARAPKKEHTVNGYVNRSLPGLIGCHRXSVYVHWVTCMGITPEVESIFLEGHTFLVRNH	300						
DB	241	AASARAPKKEHTVNGYVNRSLPGLIGCHRXSVYVHWVTCMGITPEVESIFLEGHTFLVRNH	300						
QY	301	QOASLEISPTITLTAQTLMDLGGFLISCHISSHQHGMAYVVDSCPEPOLMKKNE	360						
DB	301	QOASLEISPTITLTAQTLMDLGGFLISCHISSHQHGMAYVVDSCPEPOLMKKNE	360						
QY	361	EADYDDDLTSEMDVVRFDNDSPTQIRSAVKHPKTKWVHYIAAEEDDWDYAPLVLA	420						
DB	361	EADYDDDLTSEMDVVRFDNDSPTQIRSAVKHPKTKWVHYIAAEEDDWDYAPLVLA	420						
QY	421	PDDRSYKQYLNNQPRIGRKYKVRNMYTDTFTKREAIQHSGLIGPLLYGEVGDTL	480						
DB	421	PDDRSYKQYLNNQPRIGRKYKVRNMYTDTFTKREAIQHSGLIGPLLYGEVGDTL	480						
QY	481	LIIFPNQASRYNTPYHGTIDVRPLSRPLPKGVKHLKDFPILGELPKYKWTVVEDGP	540						
DB	481	LIIFPNQASRYNTPYHGTIDVRPLSRPLPKGVKHLKDFPILGELPKYKWTVVEDGP	540						
QY	541	TKSDPRCLTRYSSVFNMERDLASGLIGPLIICYKESVDQGNQIMSDKNVILFSVFE	600						
DB	541	TKSDPRCLTRYSSVFNMERDLASGLIGPLIICYKESVDQGNQIMSDKNVILFSVFE	600						
QY	601	NRSWYLTENIORFLPNPAGVQLEDPEFOASNMHISINGVFDLSQLSVCLHEVAYWYILS	660						
DB	601	NRSWYLTENIORFLPNPAGVQLEDPEFOASNMHISINGVFDLSQLSVCLHEVAYWYILS	660						
QY	661	IGAOTDELVSFPGSYTFKHKVYEDTLFPFSGETVFMSPENPGIWLIGCHNSDFNRG	720						
DB	661	IGAOTDELVSFPGSYTFKHKVYEDTLFPFSGETVFMSPENPGIWLIGCHNSDFNRG	720						
QY	721	MTALLKVSCKNTGDYEDSYEDISAYLLSKNAIEPSPQNSRHPSTROKQNAITI	780						
DB	721	MTALLKVSCKNTGDYEDSYEDISAYLLSKNAIEPSPQNSRHPSTROKQNAITI	780						
QY	781	PENDIEKTDWFAHRTMPKIQNVSSDILLMLLRQSPTHGLSLDQAKAYETSDDES	840						
DB	781	PENDIEKTDWFAHRTMPKIQNVSSDILLMLLRQSPTHGLSLDQAKAYETSDDES	840						
QY	841	PGAI DENNSISEMTHFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATKELDKFVSST	900						
DB	841	PGAI DENNSISEMTHFRPOLHSGDMVFTPEGLQLRLNEKLGTTAATKELDKFVSST	900						
QY	901	SNKLISTIPSDNLAAGTNTSSLGPPSPVHYDQDITTLFGKSSPLTESGGPLSLSEE	960						
DB	901	SNKLISTIPSDNLAAGTNTSSLGPPSPVHYDQDITTLFGKSSPLTESGGPLSLSEE	960						
QY	961	NNDKLLIESGLMNSQSSSWGNVSSTESGELFKGKAHGPALLTKONALFKVVISLTKTN	1020						
DB	961	NNDKLLIESGLMNSQSSSWGNVSSTESGELFKGKAHGPALLTKONALFKVVISLTKTN	1020						
QY	1021	KTSNNSATNRKTHIDGPSLLIENSFVWQNLILESTEFKKVTPLIHDRMLMDKNAIALRL	1080						
DB	1021	KTSNNSATNRKTHIDGPSLLIENSFVWQNLILESTEFKKVTPLIHDRMLMDKNAIALRL	1080						
QY	1081	NKMSNKTSSKXMEVYQKKEGPIPPDAQNPMSFPKMLFLPESARWIORTHGKNSLNSG	1140						
DB	1081	NKMSNKTSSKXMEVYQKKEGPIPPDAQNPMSFPKMLFLPESARWIORTHGKNSLNSG	1140						
QY	1141	QCPSPKQVLSLGPESVBEQNFSLSEKXKVVGKGEFTKDVGLKEMVFPSSRNJFTNLDN	1200						
DB	1141	QCPSPKQVLSLGPESVBEQNFSLSEKXKVVGKGEFTKDVGLKEMVFPSSRNJFTNLDN	1200						
QY	1201	LEENNTNQEKKIOEBIEKKFWLIQBNVVLPOIHVTGCTKNPKMLFLLSTRQNVGSGYD	1260						
DB	1201	LEENNTNQEKKIOEBIEKKFWLIQBNVVLPOIHVTGCTKNPKMLFLLSTRQNVGSGYD	1260						
QY	1261	GAYAPVLDQFRSLNDSTNRTKHTAHFSKKGBEENLEGLNCTQKIVEKYACTRISPNT	1320						
DB	1261	GAYAPVLDQFRSLNDSTNRTKHTAHFSKKGBEENLEGLNCTQKIVEKYACTRISPNT	1320						
QY	1321	SOQNFVQSRKALKQFLPLEBETELEKRIIVDDTSTQMSKMMKHTPSTLTQIDYNEKE	1380						
DB	1321	SOQNFVQSRKALKQFLPLEBETELEKRIIVDDTSTQMSKMMKHTPSTLTQIDYNEKE	1380						
QY	1381	KGAIQSPDSDCLTRSHSIPQANRSPPLIAKVSFSPSIRPIYILTVLFDQNSHLPAAASY	1440						
DB	1381	KGAIQSPDSDCLTRSHSIPQANRSPPLIAKVSFSPSIRPIYILTVLFDQNSHLPAAASY	1440						
QY	1441	RKXDSGVCSHSHFLOGAKNNLSAILTLEMTQDQREVGLGTSAKNSVTVYKVENVLP	1500						
DB	1441	RKXDSGVCSHSHFLOGAKNNLSAILTLEMTQDQREVGLGTSAKNSVTVYKVENVLP	1500						
QY	1501	KPDLPKTSGKVELLPKVHIYQKOLFPPTETNSGPHLDLVEGSLLOGTGEGAIKNEANRP	1560						
DB	1501	KPDLPKTSGKVELLPKVHIYQKOLFPPTETNSGPHLDLVEGSLLOGTGEGAIKNEANRP	1560						
QY	1561	GKVPFLVATESSAKTSPKLLDPLANDHNYGTQIPKEWKSQESPKTAFKKKDTILSL	1620						
DB	1561	GKVPFLVATESSAKTSPKLLDPLANDHNYGTQIPKEWKSQESPKTAFKKKDTILSL	1620						
QY	1621	NACSNHAIAINNGONKPEIEVTWAKQGRTERLCQNPVVLKSHOREITRTTLOSQEE	1680						
DB	1621	NACSNHAIAINNGONKPEIEVTWAKQGRTERLCQNPVVLKSHOREITRTTLOSQEE	1680						
QY	1681	IDYDDTISVEMKKEDEFDIYDEDENSQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR	1740						
DB	1681	IDYDDTISVEMKKEDEFDIYDEDENSQSPRSFOKTRHYFIAAVERLWDYGMSSSPHVLNR	1740						
QY	1741	AQSGSVPOFKVVFQEFDTGSGFTQPLVGBELNEHLGILGPVIRAEVEDNIMVTFRNQASR	1800						
DB	1741	AQSGSVPOFKVVFQEFDTGSGFTQPLVGBELNEHLGILGPVIRAEVEDNIMVTFRNQASR	1800						
QY	1801	PYSFYSSLI SYEEDORQAEPRKQNVKNETKTYEWKQHEMAPTKDFCKAWAYFSDV	1860						
DB	1801	PYSFYSSLI SYEEDORQAEPRKQNVKNETKTYEWKQHEMAPTKDFCKAWAYFSDV	1860						
QY	1861	DLEKDVHSGLGPLLVCHTNTLNPAGHQVTVQEPALFTTFDETKSWVFTEENMERNORA	1920						
DB	1861	DLEKDVHSGLGPLLVCHTNTLNPAGHQVTVQEPALFTTFDETKSWVFTEENMERNORA	1920						
QY	1921	PCNIQMEDPTFKENYRFAHNGYIMDTLPGVMAQDQIRWYLLSGMGNENIHSIFSGH	1980						
DB	1921	PCNIQMEDPTFKENYRFAHNGYIMDTLPGVMAQDQIRWYLLSGMGNENIHSIFSGH	1980						
QY	1981	VFTVAKKBEYKMALNLYPGVFETVEMLPSSKAGIWRVECLIGEHLHAGNSTLFLVSNKC	2040						
DB	1981	VFTVAKKBEYKMALNLYPGVFETVEMLPSSKAGIWRVECLIGEHLHAGNSTLFLVSNKC	2040						
QY	2041	QTPFLNGASHIHDPOITAGGQGWAPKLARLHYGSSINAKSTKBPSPKIVYDLPAPMII	2100						
DB	2041	QTPFLNGASHIHDPOITAGGQGWAPKLARLHYGSSINAKSTKBPSPKIVYDLPAPMII	2100						
QY	2101	HGKITQAGQKSSSLYSQFIIMYSLDGGKQWQIRGNSTGTILMVFVGNVDSGKIHENIFN	2160						
DB	2101	HGKITQAGQKSSSLYSQFIIMYSLDGGKQWQIRGNSTGTILMVFVGNVDSGKIHENIFN	2160						

121 VGVSYWKASEDEYEDQISOMEKEDDQVFPGESHTYVWQVLENGPMADEPPCLTYSMS 180
180 HVDLVKDLNSGLIGALLVCREGLAKETQTLHKFILLFAVFDGKGWHSGETONSMDQR 239
181 HVDLVKDLNSGLIGALLVCREGLSKERTQMLYQVLLFAVFDGKGWHSGETNSYTQSM 240
240 DAAAGARWPKHTVNGYVNRSLPGLIGCHRSVYWHVIGMGTTEVVASIFLEGHTELVRN 299
241 DSASARDWPKMETVNGYVNRSLPGLIGCHRSVYWHVIGMGTTEVVASIFLEGHTFVRN 300
300 HQASLETSPIPTTAOTLLMDLGOFLSCHISSHQHDGMEAYVKVDCSPEPQIMK-N 358
301 HQASLETSPIPTTAOTLLMDLGOFLSCHISSHQHDGMEAYVKVDCSPEPQIMK-N 360
359 NEEAEDYDDDDTDSMDVRRDDDNPSFQIORSVAKHPTKWTWHYIAAEBEDWDYAPLV 418
361 NEEAEDYDDDDY-SEMDVFTLDYDSSP-FIQIRSVAKYPTKWTWHYIAAEBEDWDYAPSV 418
419 LAPDRSYKSOYLNNGPORIGKVKVZFMAVDTSTETRAIOHESGILGPLLYGEVGD 478
419 PTDNNGSYKSOYLNNGPORIGKVKVZFMAVDTSTETRAIOHESGILGPLLYGEVGD 478
479 TULLIFKNOASRPYNIYPHGIITDVPYLSRLPKGVKELKOPPLIPGBIFKYKWTVTVED 538
479 TULLIFKNOASRPYNIYPHGIITDVPYLSRLPKGVKELKOPPLIPGBIFKYKWTVTVED 538
539 GPTKSDPRCLTYYSFVNMERDLASGLIGPLLI CYKESVDQKQNOIXSDXENVILPSVF 598
539 GPTKSDPRCLTYYSFINPERDLASGLIGPLLI CYKESVDQKQNMDSKENVILPSIF 598
599 DNRSWYLTENTORELPHAPAGVOLKDEPEQASINMHSINGYVDFSILQSLVCLHEVAYWYI 658
599 DNQSWYLTENTORELPHAPAGVOLKDEPEQASINMHSINGYVDFSILQSLVCLHEVAYWYI 658
659 LSGIAGTDFLSVFFSGYTPKHKVYEDTTLTLPFFSGETVFMENPGLMILGCHNSDPRN 718
659 LSVGAGTDFLSVFFSGYTPKHKVYEDTTLTLPFFSGETVFMENPGLMILGCHNSDPRK 718
719 RGMNTALKVSSCKNTGDIYEDSIEDISAYLSKNAIAPRFSQNSRPSIRQOFNAT 778
719 RGMNTALKVSSCKNTSDIYEDIEIPIQLVNNENVIDPERSFFQNTNHPNTRKKKXPOS 778
779 TTPENDIEKTDPMFAHRTMPKIQNVSSDMLMLRQS-PTPGLSISLQEAKEYTFSD 837
779 TTPKNDMEKIEPQFESIAEMLKVQSVSDMLMLRQSHTPTGLFLSDCQRAIYEAHD 838
838 DPSPGAIIDNSNLSMETHPRPQLEHSGDMVFTPEBSGLQELNKLGTATAELKKLDKVK 897
839 DHPSPAIIDNSNEGSKVTQLRPEHSHSEKI VFTPQGLQIRSNKSLETTIEVKKKLGLOV 898
898 SSTSNMLI-STIPSDNLAAGTDNTSILGPPSPMPVHYDSQ-DTLTFLGKSSPLTESGGPLS 956
899 SLPSPNMLITLISDNLKATEKTDGSGFPDMPVHSSKLSSTTAFGKAYSIVGSHVPLN 958
957 LSEBNDKSLLESGLMNSQESSGKNVSSSTESGRLEFKGKAHGFALLTKDNAIFKVSISL 1016
959 ASBENSNDNLDWLYMSQSLSLPRDNTLSIENDRLIEXERFHGIALITKNTLPKDNVSL 1018
1017 LKNTKTSNNAATNRKTHIDGSLLIENSPSVWQN-ILBSDEPKVTPPLHDMRLMDKNA 1075
1019 MKNTKTYNHSITNEKLTESPT-SIENSTTDLQDAILKVANSEIOEVTHLIDHGLLGKNS 1077
1076 TALPLNHSNKTTSKKNEMVQOKEGPIPPDAQNDPMSPEKMLFLPESARWIORTHGKN 1135
1078 TYLNLHMLNFTTKKNDIFHRKDEDDIPOBENTIMPFSKMLFLBESSNFWFKTGN 1137
1136 SLNSGQSPKOLYSLGPEXVEGQNTFSEKKNVWVGKEFTKDVGLKEMVFPSSRLFL 1195
1138 SLNSEQHSFKQLYVLMFKYVKNQSFSEKKNVVBODGFTKIGLKDMAFPNMSIFL 1197
1196 TNLNLHNNTHNPKKTOEBIEKKEETLIOENVLPQJHTVTGKNFMKNFLILSTRQNV 1255

1198 TTLSNVHNGRHNQERNIQSEIE-KSALIEBKVVLPVHEATGSKNPLKDILILGTFRQNI 1256
1256 EGSYDGAAYAPVQDFRSLNDSNTRTKKHTAHFSK--KGESENLEGLNQTKQIVEXYACT 1313
1257 --SLIYEVHVPVQNIUTSINNSTNTVQIHMEHFFKRRKDKETNBEGLVNKTRVWKVNY-- 1311
1314 TRISPNTSQNFVQSRKALKOPRLEPLETELEKRIIVDDTSTQWKNMKHGLTPSLTQ 1373
1312 -----PSQKNITQSRKALQOFR-----STQWLKTNCSQCTCIKQ 1349
1374 IDYNEKEKGATQSPSLDCLTRSHSIPQANRSPPLIAKVSFPSPRIRPIYITRVLFOQNSS 1433
1350 IDHSKWKKFIKTSLSLDS-SVAKSTQTNSSDSHIVKTSAPP---PIDLKRSPFQNKFS 1405
1434 HLPAASY-----RKXDSGVQESSHFOGAKXNLSLAILTLEMTGDQREVGSLGHSATNSV 1489
1406 HVQASSYIYDFKTKSSRIQESNNFLKTKINNPISLAILPWNMPIDQGFKTSFGKSNNSV 1465
1490 TYKKVENFVLPKPDLPKTSKGVVLLPKVHIYQKDLFPTETSNGSPGHLDLVEGSLLOQTE 1549
1466 TYKKRENIIFLKPILPEESGKIELLPQVSIQEBILPFTBSHSPGHLNLMKREVLQXIQ 1525
1550 GAIKWNZANRPGKVPFLRVATESAKTPSKLLDPLANDNHYGTQTPKDEWKSQBSPEKT 1609
1526 GPTKWNKAKRGE--SIKGTESSENTRSKLNHHANDYHAAQIPKDMWKSKEKSPBII 1583
1610 AFKKKDTILSNACESNHAIAMNEGONKPEIENVMAQKQTERLCSQNPVLPKHOREI 1669
1584 SIKQEDTILSRPFGNSHISGA-NEKQNPQRETTWVQGTQRTCSQIPVLPKHOREL 1642
1670 TRTILQSOQEBIDYDDTISVEMKEDFDIYDEENQSPRSFOCKTRHYFLAAVEHLWDYG 1729
1643 --SAPQSEQATDYDAITIE--TIEDFOIYSEDIKQGRSFOCKTRHYFLAAVEHLWDYG 1699
1730 MSSSPHVLNRNAQSGSVFQKVVFEFTDGSFTQFLYRGELNEHLGLLGPVIRAEVDN 1789
1700 MSTS-HVLNRVQSDNVQFKVFEFTDGSFSQFLYRGELNEHLGLLGPVIRAEVDN 1758
1790 IMVTFRNOASRPYSFYSSLSIYEBEORQGAEPKPNVPNETKTYFWKVOHHMATEDEF 1849
1759 IMVTFKNOASRPYSFYSSLSIYKEDQR-GEZPRNFVKNETKIYFWKVOHHMATEDEF 1817
1850 DCKAWAYFSDVDLEKDVHSGLLIGLILVCHTNTLNPAHGSQVTVQBFALPFTTFDETKSWY 1909
1818 DCKAWAYFSDVDLERDMHSGLLIGLILVCHTNTLNPAHGSQVTVQBFALPFTTFDETKSWY 1877
1910 FTENMERNCRAPQNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQDQIRVYLLSMGN 1969
1878 FTENVKRNCCTFCNPOVEDPTLKENYRPHAINGYVMDTLPGIWMADQDQIRVYLLSMGN 1937
1970 ENTHSHFSGHVPTRKKEEYKMAVNYLPGVFEIVEMLPKXAGIWEVCLIGEHLAGM 2025
1938 ENIQSHFSGHVPTRKKEEYKMAVNYLPGVFEILEMIPSRAGIWRVECLIGEHLAGM 1997
2030 STLFLVYVSKCOTPLGVASGHTRDFOITASGOYQWAPKLARLHYSGSINAWSTKEPFSW 2089
1998 STUFLVYVSKCOTPLGVASGHTRDFOITASGOYQWAPKLARLHYSGSINAWSTKEPFSW 2057
2090 IKVDLLAPMIHGIKTQGAQKFSYISQFIIMYSLDGKKQWOTYRGNSTGTTLWFPFNV 2149
2058 IKVDLLAPMIHGIKTQGAQKFSYISQFIIMYSLDGKKKLSYQGNSTGTTLWFPFNV 2117
2150 DSGIKHNIPNPIIARIYIRLHETHYSIRSLAWELMGCDLNSCMPLGWESKAIISDAQI 2209
2118 DSGIKHNIPNPIIARIYIRLHETHYSIRSLAWELMGCDLNSCMPLGWESKAIISDAQI 2177
2210 TASSYFTNMATWSPKARLHLOCRSNAPVNNPKEWLQVDFQKTKYVTGVTQGVKS 2269
2178 TASSYFTNMATWSPKARLHLOCRSNAPVNNPKEWLQVDFQKTKYVTGVTQGVKS 2237
2270 LLTSMYVKEFLISSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDDPILLIRYLRIHP 2329
2238 LFTSMYVKEFLISSQDGHQWTLFFQNGKVKVQGNQDSTFPVNSLDDPILLIRYLRIHP 2297

Wed Apr 14 15:52:00 2004

np000123-328-355-581.rsp

```

QY 2330 QSWVHQIALRMEVLCGERQDLY 2351
Db 2298 QIWEHQIALRLEILGCEAQOY 2319

RESULT 3
FAS_PIG STANDARD; PRT; 2133 AA.
AC P12263; Q95243;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CFB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Healey J.F., Lubin I.M., Lollar P.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 705-1573 FROM N.A.
RX MEDLINE=86287369; PubMed=3016730;
RA Toole J.J., Pittman D.D., Orr E.C., Murtha P., Wasley L.C.,
RA Kaufman R.J.;
RT "A large region (approximately equal to 95 kDa) of human factor VIII
RL is dispensable for in vitro procoagulant activity."
RN [3]
RP Proc. Natl. Acad. Sci. U.S.A. 83:5939-5942(1986).
RN [3]
RP SEQUENCE OF 392-759 FROM N.A.
RX MEDLINE=94179260; PubMed=7510693;
RA Lubin I.M., Healey J.F., Scandella D., Runge M.S., Lollar P.;
RT "Elimination of a major inhibitor epitope in factor VIII."
RL J. Biol. Chem. 269:8639-8641(1994).
CC -1- FUNCTION: Factor VIII, along with calcium and phospholipid, acts
CC as a cofactor for factor IXa when it converts factor X to the
CC activated form, factor Xa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: Contains 3 F5/8 type A domains.
CC -1- SIMILARITY: Contains 2 F5/8 type C domains.
CC -1- SIMILARITY: STRONG TO COAGULATION FACTOR V.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U49517; AAB06705.1; .
CC PIR; A25945; A25945.
CC PIR; T42763; T42763.
CC HSP; P00451; 1CFG.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR008972; Cupredoxin.
CC InterPro; IPR000421; FAS8_C.
CC InterPro; IPR008979; Gal_bind_like.
CC Pfam; PF003194; Cu-oxidase; 3.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SM00231; FAS8C; 2.
CC PROSITE; PS01285; FAS8C_1; 2.
CC PROSITE; PS01286; FAS8C_2; 2.
CC PROSITE; PS00022; FAS8C_3; 2.
CC PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
CC Blood coagulation; Repeat; Plasma; Acute phase; Calcium;
KW Signal; Glycoprotein; Sulfation.
KW SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2133 COAGULATION FACTOR VIII.
FT DOMAIN 20 357 F5/8 TYPE A 1.

```

```

20 199
207 357
399 730
583 730
760 1599
1495 1822
1495 1659
1569 1822
1822 1970
1975 2127
2127 392
392 760
760 1450
1450 1490
1490 737
737 738
738 742
742 173
173 547
547 1633
1633 1822
1822 1975
1975 233
233 259
259 601
601 929
929 985
985 1025
1025 1111
1111 1181
1181 1208
1208 1245
1245 1265
1265 1335
1335 1408
1408 1611
1611 1919
1919 713
713 734
734 792
792 1133
1133 1191
1191 1209
1209 1437
1437 1456
1456 1539
1539 1546
1546 2133
2133 AA; 239304 MW; 152BBA8997F570DA CRC64;

Query Match 64.9%; Score 8057; DB 1; Length 2133;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;

QY 1 MQELSTCFCLLRFCFSATRRYYLGAVELSDYMQSD-LGELPVDARPPVPKSPFF 59
Db 1 MQELSTCFCLLRFCFSATRRYYLGAVELSDYMQSD-LGELPVDARPPVPKSPFF 59
QY 60 NTSVYVYKTLFVEFTDLEFNIAKPRPPWMLGLPTTQAEVYDVTWILKNMASHPSLHA 119
Db 61 GPSVLYKTVFVEFTDQLFVSARPPWMLGLPTTQAEVYDVTWILKNMASHPSLHA 120
QY 120 VGVSYKASGAEDDDTSOREKEDDKVFGGSHYTVQVVLKNGKENGWASDPLCLTYSYLS 179
Db 121 VGVSWKSSGEAEYEDHTSOREKEDDKVLPKGSQTVYVQVVLKNGKENGWASDPLCLTYSYLS 180
QY 180 HVDLVKOLNSGLIGALLVCREGSLAKEKTLTKLFIILLFAVFDGKSWHSETKNSLMQDR 239
Db 181 HVDLVKOLNSGLIGALLVCREGSLAKEKTLTKLFIILLFAVFDGKSWHSETKNSLMQDR 240
QY 240 DAASARAWPKCHTVNGYVNRSLPGLICCHRSVYVHVGMTTPEVVISIFLEGHTFLVRN 299
Db 241 DPAPARAPAMHTVNGYVNRSLPGLIGCHKKSVYVHVGMTSPEVHISIFLEGHTFLVRH 300

```


Wed Apr 14 15:52:00 2004

np000123-328-355-581.rsp

DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Coagulation factor V precursor (Activated protein C cofactor).
 DE F5.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT MET-1764.
 RX MEDLINE=92232668; PubMed=1567832;
 RA Crispe L.D., Moore K.D., Kane W.H.;
 RT "Structure of the gene for human coagulation factor V";
 RL Biochemistry 31:3777-3785(1992).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT GJU-925.
 RX MEDLINE=87260886; PubMed=3110773;
 RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 RN [3]
 RP SEQUENCE OF 1-1600 FROM N.A., AND VARIANTS GJU-925 AND ILE-1285.
 RX MEDLINE=88107560; PubMed=2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 RT region of human factor V, a blood coagulation factor with four types
 RT of internal repeats";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RP SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE=86313665; PubMed=3032220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 RN [5]
 RP PARTIAL SEQUENCE FROM N.A.
 RX TISSUE=Fibroblast;
 RC MEDLINE=93203619; PubMed=8454869;
 RA Shen N.L.L., Fan S.-T., Pyati J., Graff R., Iapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes";
 RL J. Immunol. 150:2992-3001(1993).
 RN [6]
 RP SULFATION.
 RX MEDLINE=94264012; PubMed=8204629;
 RA Pittman D.D., Tomkinson K.N., Michnick D., Selighsohn U.,
 RA Kaufman R.J.;
 RT "Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full procoagulant activity.";
 RL Biochemistry 33:6952-6959(1994).
 RN [7]
 RP SULFATION.
 RX MEDLINE=90366699; PubMed=2168225;
 RA Horton G.B.;
 RT "Sulfation of tyrosine residues in coagulation factor V";
 RL Blood 76:946-952(1990).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2055-2224.
 RX MEDLINE=20052169; PubMed=1058886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V";
 RL Nature 402:434-439(1999).
 RN [9]
 RP VARIANT MET-1764.
 RX MEDLINE=95179346; PubMed=7874144;
 RA Bayston T.A., Ireland H., Olds R.J., Thein S.L., Lane D.A.;

RT "A polymorphism in the human coagulation factor V gene.";
 RL Hum. Mol. Genet. 3:2085-2085(1994).
 RN [10]
 RP VARIANT APCR GLN-534.
 RX MEDLINE=94217810; PubMed=8164741;
 RA Bertina R.M., Koeleman B.P.C., Koster T., Rosendaal P.R.,
 RA Durven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C";
 RL Nature 369:64-67(1994).
 RN [11]
 RP VARIANTS ILE-1285 AND ARG-1327.
 RX MEDLINE=96351768; PubMed=8713778;
 RA Lunghi B., Iacoviello L., Gemmati D., Dilasio M.G., Castoldi E.,
 RA Pinotti M., Castaman G., Redaelli R., Mariani G., Marchetti G.,
 RA Bernardi F.;
 RT "Detection of new polymorphic markers in the factor V gene:
 RT association with factor V levels in plasma";
 RL Thromb. Haemost. 75:45-48(1996).
 RN [12]
 RP VARIANT APCR GLY-334, AND VARIANT LYS-513.
 RX MEDLINE=98122763; PubMed=9454741;
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese";
 RL Blood 91:1135-1139(1998).
 RN [13]
 RP VARIANT APCR THR-334.
 RX MEDLINE=98122764; PubMed=9454742;
 RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C";
 RL Blood 91:1140-1144(1998).
 RN [14]
 RP VARIANTS HIS-107; THR-413; LYS-513; SER-809; THR-817; ARG-858;
 RX ARG-865; GLU-925; GLN-1146; ALA-1530; SER-1685; VAL-1749; MET-1764;
 RP ILE-1820 AND GLY-2222, AND VARIANT APCR GLN-534.
 RX MEDLINE=99318093; PubMed=10391209;
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
 RA Lander E.S.;
 RT "Characterization of single-nucleotide polymorphisms in coding regions
 RT of human genes";
 RL Nat. Genet. 22:231-238(1999).
 RN [15]
 RP ERRATUM.
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaraman N., Nemesh J., Ziaugra L.,
 RA Friedland L., Rolfe A., Warrington J., Lipschutz R., Daley G.Q.,
 RA Lander E.S.;
 RL Nat. Genet. 23:373-373(1999).
 RN [16]
 RP VARIANT APCR HIS-2102.
 RX MEDLINE=21847288; PubMed=1858490;
 RA Schrijver I., Houissa-Kastally R., Jones C.D., Garcia K.C.,
 RA Zehnder J.L.;
 RT "Novel factor V C2-domain mutation (R2074H) in two families with
 RT factor V deficiency and bleeding";
 RL Thromb. Haemost. 87:254-259(2002).
 CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
 CC with factor Xa to activate prothrombin to thrombin.
 CC -!- SUBUNIT: Factor V is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -!- DOMAIN: Domain B contains 35 x 9 AA tandem repeats, and 2 x 17 AA
 CC repeats.
 CC -!- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -!- PTM: Sulfation is required for efficient thrombin cleavage and
 CC activation and for full procoagulant activity.
 CC -!- DISEASE: Defects in F5 are the cause of Owren parahemophilia
 CC [MIM:227400], an hemorrhagic diathesis.

2113 QAKANNKOWLEIDLLKIKKIIITQOCSLSSEMVVKSYYTHYSEQGVWPKYRLKSS 2172

2299 V--KVFGNODSTPVVNSLDPLLYLRIHQSWVHQIALRMVLCGSAQDIY 2351
2173 MVDKIFEGNTKRGVKNFNPISFIRVIPKTNQSIITLRLFGC---DIY 2224

RESULT 5

FA5_PIG STANDARD; PRT; 2258 AA.
AC Q9GEP1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-NAR-2004 (Rel. 43, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
CN F5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP DOMAINS
RC TISSUE=Liver;
RX MEDLINE=21121490; PubMed=11229814;
RA Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA Kim H.K.W.;
RT "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
protein modeling of membrane binding sites and comparative anatomy of
domains."
RT Cell. Mol. Life Sci. 58:148-159 (2001).
RL C-!
CC -!- FUNCTION: Coagulation factor V is a cofactor that participates
with factor Xa to activate prothrombin to thrombin.
CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
chain, noncovalently bound. The interaction between the two chains
is calcium-dependent.
CC -!- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
and C2 may be involved in membrane binding.
CC -!- PTM: Thrombin activates factor V proteolytically to the active
cofactor, factor Va (formation of a heavy chain at the N-
terminus and a light chain at the C-terminus).
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF191308; AAC28381.1; -.
DR HSPSP; P12259; ICZT
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FAS8 C.
DR InterPro; IPR008973; Gal bind like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5 F8 type C; 2.
DR SMART; SM00231; FAS8C_2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS50022; FAS6C_3; 2.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
Repeat.
FT SIGNAL 1 22
FT CHAIN 23 2258
FT CHAIN 23 737
FT PROPEP 738 1611
FT COAGULATION FACTOR V.
FT COAGULATION FACTOR V HEAVY CHAIN (BY
FT SIMILARITY).
FT ACTIVATION PEPTIDE (CONNECTING REGION)

1198 ELSQTNLSPDLSTTLSPDLQRLN-----SALQMPISPDLSTTL 1240
1281 ---KHTAHFKGKEENLEGLGNTQKQIVKCYACTTRISNTSQNQFVQSRKALKQF 1337
1241 SPDLSTHT-----LSLDSQ-----TNLSPELSQTNL-----SPALQ- 1273
1338 RLPLETELEKRIIVDDTSTQWKNMKHLTPS-ILTOIDYNEKEKGAITQSPDLSDCL--- 1393
1274 -MPLSPDLSTTLSDFSQTNLSPDLSTHTLSPDLSTNLSP-----ALQMPISPDLSTHT 1328
1394 TRSHIPQANSPPIAKVSPSPSRPIYLRLVLPQDNSSHLPAASYRKDGSGVQESSHF 1453
1329 TSLSDFSQTNLSP-ELSQTNLSPALQWPLS-----PDSHTLSDLSQTNL----- 1373
1454 LQAKKNLSAILILEMTGTQDQREYVGLSTATNSVTYKKNVNTVLP---KPDLPKTSKGK 1510
1374 -----NLSPDLSTNLSPDLSEMPFADLSQILPTDQMTLSPDLGETLSPNFGQ 1426
1511 VELLPKV-----HIYKDLIPPTTSNGSPGHLDIV-----EGSLQGTGEGALKNE 1556
1427 KSLSPDLQSVTLSPDISQTNLSPDLSTQISPPPLDQIFYPSSSSQSLLLQ-----EENE 1480
1557 ANRPGKVPRLVATSSAKTP-----SKLDPPLAW-----DNHYGTQIPKEEKWQOE 1603
1481 S-----FPYDLCQMPSPSPILNDTFLSKENPLVIVGLSKDGTDIHILPKKEVQSSE 1535
1604 KGPETAFAKKKOTILSLNACSNHAIANEQONKPEIEVTWAKQGRTERLCSQPPVLK 1663
1536 -----
1719 IAAVERLDWYGMSSPHVLNRN--AOSGSVPQ---PKVVFQEFDTGFTQPLLYRGLNE 1773
1584 IAAEBSLDWY---SEFVQRETDIEDSDIPEDTTKVVKVFRYLDSTPTKRPGRGEYE 1639
1774 HGLGLGPIYRAEVEDNIMVTRNQARPYSFYSLLSYB-----EQRQAEPKRF 1625
1640 HLGILGPIYRAEVDDVIQRFKRLAGRPYSLHAGLSYEKSSGKTYEDDPSPEWFKEDNA 1699
1826 VKNEKTYTWKQHHMAYPKDFCDKAWAYFSDVLEKDVHSLGLPLLVCHTNTLNPA 1885
1700 VQPNSSYTYVWHATERGSPSACAWAYYSAVNEKDIHSLGLPLILCQKGLHKD 1759
1886 HGRQVTVQEPALFPTIPDETCKSYFTENMERNCRAPCNIQMEDPTFKENYRFAINGIYM 1945
1760 SNMPVDNREFVLFWMTDEKKSUYEKKRSRW-----LTSSEMKSHFEHAINGIY 1813
1946 DTIPGLMAQQRIRWILLSGNSNENTSHFSGHVFTVKKEBYKVALNLYPGVPETV 2005
1814 -SLPGLKMYEGEWRLHLNLTGGSQDTHVVFHGHCTLENGKNKHQGLGVWPLLPSSKTL 1872
2006 BMLTSGKAGIWEVEGLIHHAGMSTLFLVYSNKCOTPLGMSGHIRDFTITAGSQYQW 2065
1873 BMKASKGFWLNLTEVENQRAQMTQPLINDRCRPMGLSTGLISDSQIKASEFLGYW 1932
2066 APKLARLHYSGINAWSTKE---PFS-----WIKVDLLAPMIHGIKTQCAROKFSSLSISQ 2119
1933 EPRLARLNNNGSYNAWSVEKLAEPFASKPWLQVDMQKEVITIGITQCAKHYLKSCTYTE 1992
2120 FIIMYSLDGKKWQYRGNSTCTLWVFFGNVDSGGIKHNIENPPIIARVIRLHPHYISRS 2179
1993 FVAYSSNQINWQIFKGNSTRNVMVFNCSASTIKENQCFPIIVARIRISPTFAYNRP 2052
2180 TLMELMGCGLNCSMPLGMSKALSDAQITASSYFTNNFAT-MSPSXARLHLQGRSNW 2238
2053 TLRLSQCEVNGCSTPLQENCKTENKQITASSFKSWGWDYWFPPFARLNAQGRVNAW 2112
2239 RPQVNPKEWLOQVDFQKTMKVTVGVTTQGVKSLTSMYKVEFLISSQDGHQWTLFPQNGK 2298

CHAIN		1612	2258	(BY SIMILARITY).	CONJUGATION FACTOR V LIGHT CHAIN (BY SIMILARITY).
FT	DOMAIN	30	329	FS/8 TYPE A 1.	
FT	DOMAIN	30	193	PLASTOCYANIN-LIKE 1.	
FT	DOMAIN	203	329	PLASTOCYANIN-LIKE 2.	
FT	DOMAIN	348	683	FS/8 TYPE A 2.	
FT	DOMAIN	348	525	PLASTOCYANIN-LIKE 3.	
FT	DOMAIN	535	683	PLASTOCYANIN-LIKE 4.	
FT	DOMAIN	691	1611	B.	
FT	DOMAIN	1168	1539	41 X 9 AA APPROXIMATE TANDEM REPEATS OF T-L-S-P-D-L-[GS]-[HQ]-T.	
FT	REPEAT	1168	1176	1.	
FT	REPEAT	1177	1185	2.	
FT	REPEAT	1186	1194	3.	
FT	REPEAT	1195	1203	4.	
FT	REPEAT	1204	1212	5.	
FT	REPEAT	1213	1221	6.	
FT	REPEAT	1222	1230	7.	
FT	REPEAT	1231	1239	8.	
FT	REPEAT	1240	1248	9.	
FT	REPEAT	1249	1257	10.	
FT	REPEAT	1258	1266	11.	
FT	REPEAT	1267	1275	12.	
FT	REPEAT	1276	1284	13.	
FT	REPEAT	1285	1293	14.	
FT	REPEAT	1294	1302	15.	
FT	REPEAT	1303	1311	16.	
FT	REPEAT	1312	1320	17.	
FT	REPEAT	1321	1329	18.	
FT	REPEAT	1330	1338	19.	
FT	REPEAT	1339	1347	20.	
FT	REPEAT	1348	1356	21.	
FT	REPEAT	1357	1365	22.	
FT	REPEAT	1366	1374	23.	
FT	REPEAT	1375	1383	24.	
FT	REPEAT	1384	1392	25.	
FT	REPEAT	1393	1401	26.	
FT	REPEAT	1402	1410	27.	
FT	REPEAT	1411	1419	28.	
FT	REPEAT	1420	1428	29.	
FT	REPEAT	1429	1437	30.	
FT	REPEAT	1438	1446	31.	
FT	REPEAT	1447	1455	32.	
FT	REPEAT	1456	1464	33.	
FT	REPEAT	1465	1473	34.	
FT	REPEAT	1474	1482	35.	
FT	REPEAT	1483	1491	36.	
FT	REPEAT	1492	1500	37.	
FT	REPEAT	1501	1509	38.	
FT	REPEAT	1510	1518	39.	
FT	REPEAT	1519	1527	40.	
FT	REPEAT	1531	1539	FS/8 TYPE A 3.	
FT	DOMAIN	1616	1941	PLASTOCYANIN-LIKE 5.	
FT	DOMAIN	1616	1785	PLASTOCYANIN-LIKE 6.	
FT	DOMAIN	1795	1941	FS/8 TYPE C 1.	
FT	DOMAIN	1941	2095	FS/8 TYPE C 2.	
FT	DOMAIN	2100	2255	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	SITE	737	738	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	SITE	1029	1030	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	SITE	1611	1612	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).	
FT	DISULFID	167	193	BY SIMILARITY.	
FT	DISULFID	248	329	BY SIMILARITY.	
FT	DISULFID	499	525	BY SIMILARITY.	
FT	DISULFID	602	683	BY SIMILARITY.	
FT	DISULFID	1759	1785	PROBABLE.	
FT	DISULFID	1941	2095	BY SIMILARITY.	
FT	DISULFID	2100	2255	BY SIMILARITY.	
FT	MOD_RES	692	696	SULFATION (POTENTIAL).	
FT	MOD_RES	696	696	SULFATION (POTENTIAL).	

FT	MOD_RES	724	724	SULFATION (POTENTIAL).	
FT	MOD_RES	726	726	SULFATION (POTENTIAL).	
FT	MOD_RES	745	745	SULFATION (POTENTIAL).	
FT	MOD_RES	1560	1560	SULFATION (POTENTIAL).	
FT	MOD_RES	1576	1576	SULFATION (POTENTIAL).	
FT	MOD_RES	1581	1581	SULFATION (POTENTIAL).	
FT	MOD_RES	1584	1584	SULFATION (POTENTIAL).	
FT	MOD_RES	1588	1588	SULFATION (POTENTIAL).	
FT	MOD_RES	1631	1631	SULFATION (POTENTIAL).	
FT	CARBOHYD	153	153	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	225	225	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	297	297	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	382	382	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	467	467	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	553	553	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	741	741	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	752	752	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	760	760	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	776	776	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	782	782	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	899	899	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	960	960	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1048	1048	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1066	1066	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1073	1073	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1089	1089	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1174	1174	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1480	1480	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1537	1537	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1597	1597	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1737	1737	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1886	1886	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2044	2044	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	2243	2243	N-LINKED (GLCNAC. .)	(POTENTIAL).
SQ	SEQUENCE	2258	2258	AA; 256078 MW; 9159B9E0076A2ACC CRC64;	

Query Match

Best Local Similarity 30.4%; Pred. No. 4.7e-135;

Matches 768; Conservative 388; Mismatches 874; Indels 497; Gaps 77;

QY	22	RYILGAVELSWDMQSDLGELPVDARFPFRVPKSPFPNTSVVYKTLFVEITDHLFNIA	81
Db	32	RQFYVAAQSIWNTHPE-----PTHSSSPFRATS--FKIVVREY-EAYFQKE	76
QY	82	KSRPPWNGLLGPTIOAEVYDVTVITLKNMASHPVSLHVGVSYSKASAGARYDDQTSORE	141
Db	77	KPSRMSGLLGPLYADVGDMKVHFNKADKPLSIHPQGIKYSKFAEGASYPDHTFLVE	136
QY	142	KEDDKVFPGSSHTYVMQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREG	201
Db	137	KMDDAVAPQOEYIYENNISDSGTHNDPPCLTHIYYSYENLIQDFNSGLIGLLICKKG	196
QY	202	SLAKEKOTIL--HKFILLFAVDEGKSWHSETKNSLNQDRDAASARAWPQOHVWNGVNR	259
Db	197	LTLEDGIQKMFQYVIMFAVDFDESKSWQSSS-----LMYVNGVNG	240
QY	260	SLPGLIGCHREKSVYVHVGITTPVEHSIFLEGHTFLVRNHRQASLEISPTILTATLL	319
Db	241	TMDITVCAYDHSIHLWHLIGNSSGPEFSIHFSQVLEQNHKKVSATLTVSATTTANMTV	300
QY	320	MDLQGLLSCHISHQHDGMEAYVVDSCPEEPQLINKNEEAEDDDDDTJDEMDVVR	379
Db	301	SPEGKWFISLLIKHFOAGQCAVIDIKNCARTRKPKK-----LTRDQ-----	343
QY	380	DDNSPSFIQIRSVAKGPKTWVHYIAABEEDVDYAPLVLPDDRYSKYSLNNGPORIG	439
Db	344	-----RRHIXWEYFTAABEVIWDYAPILIFANMDKKYRSHLDNFSNQIG	388
QY	440	RKYKVFMAVTTDETFR--EATQCHESGILGPLLYGEVGDTLIIIFKQASFPYNIYPH	497

Db 389 KHYKVVYKQDESFTKRLNPNKEDGILGPVIRAQVDTLKVFNKMASRYSIYPH 448
 Qy 498 GIT-----DVRPYSRRLPKGVXHLKDFPILPGEIPIKYKWTVVVBGPXTKSDPRCLTRY 551
 Db 449 GVTSPYEDDVNSSTSDNMTIR-----AVQPGETVYKWNILESDPTENDAOCLTRP 503
 Qy 552 YSSFNMRDLASGLIGLPLCYKESVDOKNOJMSDKRNVLSPFEDENRSWYLTENIQ 611
 Db 504 YYSNVDTIEDIASGLIGLPLCKRSLDKRGLOQTADIEQKAVFAVDENKSWIEDNIY 563
 Qy 612 RFLPNAGVOLDEPFOASNMHSINGVYFDSL-OLSVCLIEHAYVYILS-GAQTDFLSV 670
 Db 564 KPCNPEKVKDDPKFYESNMSTINGVYFESIPTLGFCFDDTVQWHFCSVRTHNULTI 623
 Qy 671 PFSGYTFKXMYEJTLPLFPSPGTFVMSNENPGLWILGCHNSDFRNGMTALLKVSSC 730
 Db 624 HPTGHSFYKGRHEDTLPLFMKESGVTVMNVTMLTMTNSPNRKKQLKFRDVKC 683
 Qy 731 DMTG-DYVEDSYEISAYLL-----SKNAIEP-----RSFQNS 765
 Db 684 IRDDDESYEIIYEPSSSTLTKRKHDSSEKKEENDEYDQDLASVLGIRFSNS 743
 Qy 766 REPSTKQFNATTIP-ENDIEKTDPFWAHRCFMPKIQNVSSDMLMLRQSPTHGLSL 824
 Db 744 LY-QEDDEFNLTALENNSEFIPP-----STD-----RAVDSNSSP 780
 Qy 825 SLOBAKYTFDSD-----PSPGADSNNS-----LSEMTHERPOLHSGDMVFTE-BSG 873
 Db 781 GNISRAPANTFEPKXILHPKATKAGSPRHTGLVKNLVLNRRRQHSOPSEDIENP 840
 Qy 874 IQ-----LALNE-----KLGTAATE-LKLDKFKVSTSNLSTIP 909
 Db 841 LOSVITGILLSPTGTGRNREHPKHKRPFAGDQAKHFPQMEPPAKTGRH-----IS 896
 Qy 910 SDMLAAGTONTSLGPPSPVHYDSOLDITLPGKSSPTESGGLPSLSENN-----962
 Db 897 QDN-----SSSSMGPLE-----DLSSDILLERK-DPSTINGKWLHVSEKGEIYQDA 945
 Qy 963 DSKLLEGLMNSOBS-----SWGNKVSSTESGELFKGKRAHGPALLTKDINALFKVISILLKT 1019
 Db 946 DEDVAVNKLFPNNPONASRGENIIPFNK-----HGKQRGHPIFVTRH-----KLQE 993
 Qy 1020 NKTSNNS-----ATNKTHIDGSLIENSPPVQWQIL-ESDTEFKKVTPLIHDM 1069
 Db 994 RQDEGNILKKGRFFITRKKRKKKRVHVPSPSPENPLRGEANTFSS-----1043
 Qy 1070 LMDXNATALRNLHNSKNTSTKNNMVQKKEGIPDPAQH-----PDMSFFKMLFPPSAR 1126
 Db 1044 --DRRQNSHLLHESNET-----FPPTDLNQTFPSMNLSLIASHPD---1082
 Qy 1127 WQORTHGKNSLN-SQOGSPKQVLSLGPESVQGNFLSEKKNVVGKGEFTKQVGLKM 1185
 Db 1093 -----HNQNLNDTHOTSPP-----LDLYQT 1103
 Qy 1186 VPPSS--RNLFLTLNLDNLHNNHTNQEKKIOEIEKKEKTLIOENVVLPOHTVTKNFM 1243
 Db 1104 VTPDEPYTAPLODLDPHSTAVPSHQSSLPEPI-----QME-----1140
 Qy 1244 KNLFLILLSTRQNVGSDYDGAIVPVLQDFRSLNDSNTRTKHTAHP--KKGEENLEG 1300
 Db 1141 -----DYD-----LRNKASFTPVSENFSSFLKLAGHRTSPDL-1173
 Qy 1301 NOTKQIVKVIYACT-----TISNTSQNFVQTSKRALQPRILPEETEL-----EK 1348
 Db 1174 NOTLSPELSQTTLSDPGHVTLSPDLSTQTTLSPLDLSHTLSP---DLGHTTSLDLSHT 1230
 Qy 1349 RIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITQSPDLSDCLTRSHSIFQANRSP 1403
 Db 1231 TSLSPLSQTTLSPLDLSHTLS-----PDLGHTTSLSPDLSTLSPDLGHATLSP-D 1280
 Qy 1409 IAKVSFPSPIRPTVTRVLFQNSHLPAASTRKXDSG-VQESSHFLOAKKNNLSLATL 1467
 Db 1281 LSHVTLSPDLGHTTSLSPDLGHTTSLSPDLGHTTSLSSDVSHTTSLSPDLSTLSPDL 1340

Qy 1468 T--LEMTGDQREYSGISGTSATNSVTYKXVENTVLP-----KPDLPKTS-----GKVEL 1513
 Db 1341 SPDLSHHTLSPDLGHTTSLSPDLSTQTTLSPLDGHMTLSPDLRQTSHPDPDKTSY--1502
 Qy 1514 LPKVHIYQKDLPTFETNSGPGHLDELVEGSILOQTEGAIKWNANRPGKVPFRLVATESS 1573
 Db 1395 SP-----DLSHT-TLSPDLGHTLSPD--LGOTTLSPDLGHTTSLSPDLSHMTLSSELS 1444
 Qy 1574 AKTBSKLL-----DPLAW-----DNHYGTQIPKEEWSQSKSPKTAFAKK 1613
 Db 1445 HETLSPDLSTQTTLSPLDSEIIPFSPDLAQTTLSSDLNETTSLSPDLRQTSHPDPDKTSY--1502
 Qy 1614 KDTLSLNACSNHAIANAINEGQNKPEIEVWAKQGERLCSQ-----NPVLAERHQ 1666
 Db 1503 -----ISBSSQSVTLPEFGQTSPPDLGQRPSPFSSHSTLNNFTPREFNPVWVUGLS 1554
 Qy 1667 RE-----IRFTLQSDOE-----IDYDQTTISVEMKEDFDIYDENQSP-----1707
 Db 1555 RDDGYVBIIPROQSESEEDYVKIDYVEDDPYQTDVTR-----DINSRNPDAAM 1608
 Qy 1708 --RSPOKXTRHYFAAVERLWDYGMSSSPHVLNRQAQSGVPO---FKKVVFQETDGSF 1762
 Db 1609 YLRNNGNRNVYJAAEELSMYDK-----FTQREDIDVPEHTIYKXVVERKYLDSTF 1662
 Qy 1763 TOPLVRGELNHELGLLGPYIRAEVEDNTMVTRNQAASRPSFYSSLSIYE-----ED 1814
 Db 1663 TKLDRPGEVEEHLGILGPILRAEVDVIOVRPKNLASRPSLHAGLSYEKSSSEKTYED 1722
 Qy 1815 QPQGAEPKPNFKVKNETKYFWKVQHVMAPTDEFCKAWAYFSDVLEKDVHSLGIGPL 1874
 Db 1723 DSPEWFKEDNAVQPNSSYIYWHATERSGPSPGACAWAYSAVNPEKDTHSLGIGPL 1782
 Qy 1875 LVCHTNTLNPAAHGRQVTVQEPALFTTIPDETYSWVFTENMERNCRAPCMIQMEDTFKEN 1934
 Db 1783 LICKCKTLHKENNPVDMREFEVLLFMVFEKKSYYEKKFTRSW-----LTSSEVKNS 1836
 Qy 1935 YRPHANGVIMDT--PGLVMAQOORIRWILLSMGSNENIHSIHFSGHVFTVRKKEEYKMA 1994
 Db 1837 HKPHAINGMYN-LPGLRMYEOEWVRLHLLNLGGSDIHHVHPHGOTLLENGTQOQHQLGV 1895
 Qy 1995 YNLYPGVEETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGNASGHIRDP 2054
 Db 1896 WPLLPGSPKTLNKTSKAGWLLDTEVGENQORAGMOTPFLIIDRECKMPMGLSTGLIADS 1955
 Qy 2055 QITASQYQWAPKGLARLHYSGSINAWSTKEPS-----WIKVDLLAEMIIHGKTOG 2107
 Db 1956 QIKASEFWGHQPKLARLNNGSYNAWIT-DKPSGESNSKPMIQQVDMQREVVFTGIGTQG 2014
 Qy 2108 ARQKFSLSYTSOFLIMVSLQKXQWYRGNSTGLMVFFGNVDSGIGIKHINFPPIIARY 2167
 Db 2015 AKYLLKSYTYTFEFNAVSSQORNRWIFKGNSTKNWYFNGNSDASTITENQFDPVVARY 2074
 Qy 2168 IRLPHYTSIRSTILRMEIMGCDLNSCMPLGMSKASIDAOITASSYFTNNFAT-NSPSK 2226
 Db 2075 IRISPTSYNKPALRLELQCEVNGCSTPLGMSGNIKNEQITASSFKSWGWDYWEFFR 2134
 Qy 2227 ARLEHQRSNARPOVNPXENLQVDFQKTMKVTVTQGVKSLTSMYKKEFLISSQD 2286
 Db 2135 ARLNAQGRVNAQWAKANNNNQWLQIDLLKTKITAITTQCGCKSLUSSEMYVKRYTIOVSDR 2194
 Qy 2287 GHOWTLFPQNGKV--KYFQGNQDSFTPVVNSLDPPLLTRYLIRHPQSWVHOIALRMEYLG 2344
 Db 2195 GVEKRSYREKSSWVKDFEGNNNKGHVKNFFNFPIISRTIIRIIPKWNQSIARLRLBLFG 2254
 Qy 2345 CEAOOLY 2351
 Db 2255 C---DIY 2258

AC	Q28107; Q28108;	FT	DOMAIN	30	327	F5/8 TYPE A 1.
DT	01-NOV-1997 (Rel. 35, Created)	FT	DOMAIN	30	193	PLASTOCYANIN-LIKE 1.
DT	01-NOV-1997 (Rel. 35, Last sequence update)	FT	DOMAIN	203	327	PLASTOCYANIN-LIKE 2.
DT	15-MAR-2004 (Rel. 43, Last annotation update)	FT	DOMAIN	348	686	F5/8 TYPE A 2.
DE	Coagulation factor V precursor (Activated protein C cofactor).	FT	DOMAIN	348	525	PLASTOCYANIN-LIKE 3.
GN	F5.	FT	DOMAIN	535	686	PLASTOCYANIN-LIKE 4.
OS	Bos taurus (Bovine).	FT	DOMAIN	696	1564	B.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	FT	DOMAIN	1124	1151	2 X 14 AA TANDEM REPEATS.
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	FT	REPEAT	1124	1137	1-1.
OC	Bovidae; Bovinae; Bos.	FT	REPEAT	1138	1151	1-2.
OX	NCBI_TaxID=9913;	FT	DOMAIN	1188	1453	30 X 9 AA APPROXIMATE TANDEM REPEATS OF
RN	[L]	FT	REPEAT	1188	1196	[AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
RP	SEQUENCE FROM N.A.	FT	REPEAT	1197	1205	2-1.
RC	TISSUE=Liver;	FT	REPEAT	1206	1214	2-2.
RX	MEDLINE=92147638; PubMed=1737753;	FT	REPEAT	1215	1223	2-3.
RA	Guinto E.R., Esmen C.T., Mann K.G., Macgillivray R.T.;	FT	REPEAT	1224	1232	2-4.
RT	"The complete cDNA sequence of bovine coagulation factor V.;"	FT	REPEAT	1233	1241	2-5.
RL	J. Biol. Chem. 267:2971-2978(1992).	FT	REPEAT	1242	1250	2-6.
RL	[2]	FT	REPEAT	1251	1259	2-7.
RP	DISULFIDE BONDS.	FT	REPEAT	1260	1268	2-8.
RP	MEDLINE=95034740; PubMed=7947716;	FT	REPEAT	1269	1277	2-9.
RA	Xue J., Kalafatis M., Silveira J.R., Kung C., Mann K.G.;	FT	REPEAT	1278	1286	2-10.
RT	"Determination of the disulfide bridges in factor Va heavy chain.;"	FT	REPEAT	1287	1295	2-11.
RL	Biochemistry 33:13109-13116(1994).	FT	REPEAT	1296	1304	2-12.
CC	-!- FUNCTION: Coagulation factor V is a cofactor that participates	FT	REPEAT	1305	1313	2-13.
CC	with factor Xa to activate prothrombin to thrombin.	FT	REPEAT	1314	1322	2-14.
CC	-!- SUBUNIT: Factor Va is composed of a heavy chain and a light	FT	REPEAT	1323	1331	2-15.
CC	chain, noncovalently bound. The interaction between the two chains	FT	REPEAT	1332	1340	2-16.
CC	is calcium-dependent.	FT	REPEAT	1341	1349	2-17.
CC	-!- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14	FT	REPEAT	1350	1358	2-18.
CC	AA repeats.	FT	REPEAT	1359	1367	2-19.
CC	-!- PTM: thrombin activates factor V proteolytically to the active	FT	REPEAT	1368	1376	2-20.
CC	cofactor, factor Va (formation of a heavy chain at the N-	FT	REPEAT	1377	1385	2-21.
CC	terminus and a light chain at the C-terminus).	FT	REPEAT	1386	1394	2-22.
CC	-!- PTM: Sulfation is required for efficient thrombin cleavage and	FT	REPEAT	1395	1403	2-23.
CC	activation and for full procoagulant activity (By similarity).	FT	REPEAT	1404	1412	2-24.
CC	-!- SIMILARITY: Contains 3 F5/8 type A domains.	FT	REPEAT	1413	1421	2-25.
CC	-!- SIMILARITY: Contains 2 F5/8 type C domains.	FT	REPEAT	1422	1430	2-26.
CC	-!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.	FT	REPEAT	1431	1439	2-27.
CC	-----	FT	REPEAT	1440	1448	2-28.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	FT	REPEAT	1449	1457	2-29 (PARTIAL).
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	FT	DOMAIN	1569	1590	F5/8 TYPE A 3.
CC	the European Bioinformatics Institute. There are no restrictions on its	FT	DOMAIN	1591	1738	PLASTOCYANIN-LIKE 5.
CC	use by non-profit institutions as long as its content is in no way	FT	DOMAIN	1748	1890	PLASTOCYANIN-LIKE 6.
CC	modified and this statement is not removed. Usage by and for commercial	FT	DOMAIN	1894	2048	F5/8 TYPE C 1.
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	FT	DOMAIN	2053	2208	F5/8 TYPE C 2.
CC	or send an email to license@isb-sib.ch).	FT	SITE	741	742	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CC	-----	FT	SITE	1034	1035	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
DR	EMBL; M81440; AAA30512.1; -	FT	SITE	1564	1565	CLEAVAGE (BY THROMBIN).
DR	EMBL; M81441; AAA30513.1; -	FT	DISULFID	167	193	PROBABLE.
DR	FIR; A42580; KFE05.	FT	DISULFID	248	329	BY SIMILARITY.
DR	HSSP; P12259; 1CZT.	FT	DISULFID	499	525	BY SIMILARITY.
DR	InterPro; IPR001117; Cu-oxidase.	FT	DISULFID	607	688	SULFATION (POTENTIAL).
DR	InterPro; IPR008972; Cupredoxin.	FT	DISULFID	1712	1738	SULFATION (POTENTIAL).
DR	InterPro; IPR000421; FA58 C.	FT	DISULFID	1894	2048	SULFATION (POTENTIAL).
DR	InterPro; IPR008979; Gal_bind_like.	FT	MOD_RES	697	697	SULFATION (POTENTIAL).
DR	Pfam; PF00394; Cu-oxidase_3.	FT	MOD_RES	701	701	SULFATION (POTENTIAL).
DR	Pfam; PF00754; F5_F8_Type_C_2.	FT	MOD_RES	730	730	SULFATION (POTENTIAL).
DR	SMART; SM00231; FA58C; 2.	FT	MOD_RES	1513	1513	SULFATION (POTENTIAL).
DR	PROSITE; PS01285; FA58C_1; 2.	FT	MOD_RES	1529	1529	SULFATION (POTENTIAL).
DR	PROSITE; PS01286; FA58C_2; 2.	FT	MOD_RES	1537	1537	SULFATION (POTENTIAL).
DR	PROSITE; PS00022; FA58C_3; 2.	FT	MOD_RES	1541	1541	SULFATION (POTENTIAL).
DR	PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.	FT	CARBOHYD	225	225	N-LINKED (GLCNAC. . .)
KW	Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;	FT	CARBOHYD	239	239	N-LINKED (GLCNAC. . .)
Repeat.	SIGNAL.	FT	CARBOHYD	297	297	N-LINKED (GLCNAC. . .)
FT	CHAIN	FT	CARBOHYD	382	382	N-LINKED (GLCNAC. . .)
FT	CHAIN	FT	CARBOHYD	460	460	N-LINKED (GLCNAC. . .)
FT	CHAIN	FT	CARBOHYD	553	553	N-LINKED (GLCNAC. . .)
FT	PROPEP	FT	CARBOHYD	587	587	N-LINKED (GLCNAC. . .)
FT	CHAIN	FT	CARBOHYD	745	745	N-LINKED (GLCNAC. . .)
FT	CHAIN	FT	CARBOHYD	756	756	N-LINKED (GLCNAC. . .)

FT	CARBOHYD	774	774	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	755
FT	CARBOHYD	780	780	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	755
FT	CARBOHYD	902	902	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	737
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	84
FT	CARBOHYD	964	964	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	783
FT	CARBOHYD	1044	1044	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	783
FT	CARBOHYD	1053	1053	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	864
FT	CARBOHYD	1062	1062	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	838
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	904
FT	CARBOHYD	1078	1078	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	898
FT	CARBOHYD	1094	1094	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	954
FT	CARBOHYD	1451	1451	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	942
FT	CARBOHYD	1490	1490	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	1004
FT	CARBOHYD	1550	1550	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	997
FT	CARBOHYD	1690	1690	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	1054
FT	CARBOHYD	1839	1839	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	1048
FT	CARBOHYD	1997	1997	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	1144
FT	CARBOHYD	2196	2196	N-LINKED (GLCNAC. . .)	(POTENTIAL. . .)	1087
FT	VARIANT	587	592	NETLPA -> T (IN VARIANT 2).		1171
SQ	SEQUENCE	2211	AA; 248981 MW; CEBF90B738667C45 CRC64;			1231
Query Match 22.1%; Score 2743; DB 1; Length 2211;						
Best Local Similarity 30.2%; Pred. No. 2.4e-133;						
Matches 764; Conservative 383; Mismatches 833; Indels 550; Gaps 80;						
QY	22	RYIYLGAVELSDWYQSDLGELPVDARPPRVKSPDNTSVYKKTFLVBFSTDLHENIA	81			
DB	32	ROFYAAQSIIRNYR-----PESTHL-----SSKPFETS--FKIVIVREX-EAYFOKE	76			
QY	82	KRPFWMLGGTIOAEVYDVITLKNMASHPVSLHVGVSXKASEGAEYDDTOSORE	141			
DB	77	KFQRTSGLLGTLTAEVGDIMKVFHKNKAKHPLSLHAQGIKVSXKFECASTDTLPM	136			
QY	142	KEDDKVFPQGSITYVQVLKENGPMASDPLCTYSYLVSHVDLVKOLNSGLIGALLVCREG	201			
DB	137	KMDAVAPQOEYTYEIISEHSGPTDDPPCLTHIYISVNLVEDFNSGLIGPLLCRKG	196			
QY	202	SLAKEKOTL--HKFTLLFAVDEGSHSEFKNSLUMQDRDAASARAWPKMHTVNGYVR	259			
DB	197	LTEDGTQKFXEQHVLMAVDFDESXWNTSS-----LMTVNGYVNG	240			
QY	260	SLPGLIGCHRSYVHVIGMTTPEVHSIFLEGHTEFLVNRHQASLEISPTIFLTATLL	319			
DB	241	TWPDITVCAHDHLSWHLIGMSGPELFSIHFNQVLEQNHKISAITLVSATSTTANMV	300			
QY	320	MDLQGLLSCHSSHQHDGMEAVKVDSCBEPBOLIMKNEBREDDDDLTJSEMVMVRF	379			
DB	301	SPEGRWTIASLIPRHFQAGMAYIDIKNCAKKTENPKK-----LTRDQ-----	343			
QY	380	DDNSFSFIQIRSVAKGHPKTWVHYIAAEBEDWDYAPLAPDDRYSYKSOYLNNGPQRI	439			
DB	344	-----RAHRIKWEYFIAAEVINDYAPIIPANNVKKVRSIHLNDFSNRIG	388			
QY	440	RKYKVRXYAYTDETTKTR--EAIQHESGILGELLGYEGVDTLLIPKQOASPYNIYH	497			
DB	389	KHYKVVYKQYQDSTFKRLDPSSEGDGILPIIRAQVRDTLKIIVFKNWSRSYIYPH	448			
QY	498	GIT-----DVRPLYSRLPAGVKHLDKPPILPCEIFKYKWTVVEDGPTKSDPRCL	548			
DB	449	GVTFPSYDNBNVSSSTSGSNMTYRAVR-----PGETTYTKWNILESDEPTENDAOCL	500			
QY	549	TRYSSFPVNMERDLASGLIGPLTICYKESVDQKQNOIMSKRNVILFSPVDENRSWYLTE	608			
DB	501	TRPYNSVDITRDLASGLIGLLICKESLDRGIGQAADIEQOAVFAVFDENKSWYIED	560			
QY	609	NIQFLPNPAGVQLDEDFQASNTMH-----SINGYVDFSILQ-LSVCLHEVAVHYIISIG	662			
DB	561	NIYKFCENPEXKVRDDPKFYESNMVNTLPAINGVYVPESIFILGFCFDDTVQVHFCVSG	620			
QY	663	AQTFLSVFSGYFKKMWYEDTLTLPFSGSEVFMVSMENPGLWILGCHNSDFRNGMT	722			
DB	621	TQNDILATHTGHSFYKGRHEDTLTLPFMOGESVVTMNVGVTWMLTWMNSNPRSKLR	680			

QY	723	ALLKSSCDKXTQDYVEDSYEDI-----SAYLLSKNA-----	755
DB	681	LRPRDAKTRNDOD--DSYEIIYFPGSGTAMTKKHDSSEIEDENADSDYQDEIALI	737
QY	756	IEPRFSQNSRHSSTRQKQGNATIP--ENDIEKTDWFAHRTMPKIQNVSSSDLLMLLR	84
DB	738	IGLRSFNSSL--NOEKDELNLTALALEKDESEIPP--SANKS-----LDSNSS--	783
QY	815	QSPTPHGLSLSLDQEAKEYTFSDPSPGA-----IDSNNSLSMETHFRQLHSHG----	864
DB	784	RSHVRLIAKFAESLKTLLHLEAPAGSPLHAGLDKNSALNP-----PWABHSSPYSE	838
QY	865	-----DMVFTPEGLQJ-----RLNEXLGTAAATELKLDFKVSSTSNL	904
DB	839	DPREHPLSDVTGVSLPFGTFKPKAPKQRPQVGRQAACHKFSQTRPPAKHTRL	898
QY	905	ISTIPSNLAAGTNTSS--LGPPSMPVHVDSDQDITLFGKSS-----PLTESGGP	954
DB	899	-----SQNSSSRMGP-----WEDIPSDLLLLQKQDFPKILNGEHLVSEKGS	942
QY	955	LSLSEENNDSKLESGLMNGQES--SWGKNVSSTESGRIFKGRAGHPALLT-----	1004
DB	943	YEIIQDANENKTVAK--LSPSPONDSRTWGENIPFKNS-----HGKQSGHPTPLVTRKPLQ	997
QY	1005	-----KDNALFKVS--SLUKTKNTSNNSATNRKTHIDGSLLENPSVWQ-----NILES	1054
DB	998	DRQDRNSRLKEGLFIR-----TRKKKBEKPAHVPLSPFRSFPHPURGEVNASFS	1048
QY	1055	DTEPKVKTPLTHDMMLDKNATALRNHMSNKTSSKNMVMQOKGPIFPDPAQNDMS	1144
DB	1049	DRRN-----HSLLLHASNETSLID--LNQTFPSWNLSLA-----ASLPD--	1087
QY	1115	FFKMLFELPESARWQORTHGKNSLN--SQGSPSPQLV--SLOPEKSVEQCNFLSEKNKVV	1171
DB	1088	-----HDQSPNDTTSQTSPPDLYPTVSPSEHYQ-----	1117
QY	1172	KGKFTKGVKENVFPSSRNLFNLNLNHNHNTNQEKKIQEBIEKKTLLIQENVLP	1231
DB	1118	-----IFP-----IQSDPTHTSTAPNSRSPD	1139
QY	1232	QIHTVTGKNFMKMLFLLSTRONVEGSD--GAYAVLQDFRSL-----	1273
DB	1140	PHSTTAPNSRP-----PTQSPQPNVLDNRAPITDVSQIFPSLELEVWQATSLDLS	1194
QY	1274	NDSTNRKTKHTAFKKGEBEENLEGNGOTKOIVKAYACTTRISNTSOONFVTOESKEA	1333
DB	1195	QFSIPDLQOMALSPDQBSLSPDLQ-----TSLSPDLQBSLSPDLQGTQ	1242
QY	1334	LKQFRPLPEET--ELEKRIYDDTSTQWSKNMKHLP-----STULTQIDYNEKEGAIQ	1386
DB	1243	LSPD--PSQESLSPDLQGTALSPDPS-----QESLSPDLQGTALSPDPGQESLSPDLQ	1294
QY	1387	SPLSCLTR--SHGTFQANRSPFLPIAKVSSPPSRPIYLRVLPQDNSSHUPAASRYKK	1443
DB	1295	TSLSPDLQBSLSPDLQGTALSPDP-----SQBSLSPDLQGTALSPD-----PSQBSLSP	1344
QY	1444	DSGVQSSSHFLQAKKNLSLAILTLEMTQDQEVCSLGTSAATNSVTYKVKVENTVLPKPD	1503
DB	1345	DLQOTSLSPDL--QBSLSPDLQGTALSPDPQBSL-----SPD	1381
QY	1504	LPKTS-----GKVELLPKVHIYQKDLFPPTETNSGSPGHLDLVEGSLLOQTGGAIKWNEAN	1558
DB	1382	LGQTSLSLSPDLQBSLSP-----DLQGTALSP-----DLQBSLSPDL-----	1418
QY	1559	RPKVPFLRVATESAKTFSLKDLPLAWNHYGTQIPKSEWKSQESKPEKTAFFKXDTI--	1617
DB	1419	--QOTPL-----SPDLSLESL-----SPDLSLQDLKQTS	1446
QY	1618	--LSIN-----ACENHAIANEQONPETEVTWAKQGRTERLCSQ-----NPPVLKRX	1665
DB	1447	PLDLNQTSHTSSESSQLPPEFCQTFPNADIGQMPSPDSTLNTFTIPEEFNPLVVUGL	1506
QY	1666	QRE-----ITRTLLQSQEE-----IDYDOTISVEMKKEPDIYDEENQSP-----	1707

Db 1507 SRDGGYIEIIPQKESSEDEGEFEFVAYNDPQYDRLT-----DINSRNPNDIAA 1560
 QY 1708 ---RSQKTRHYFIAAVERLWYDYGSSSHVLNRASQGS---VQO---FKVWVQEF 1758
 Db 1561 WYLRNTGNKYYVIAAEISWDYS-----KFQSDVDVYVPEDTYKWKVFRKYL 1611
 QY 1759 DG8FTQLYGELNEHGLGIPYRAEVEDNIMVTPNQASRPVSFYSSLSIYE----- 1812
 Db 1612 DSTFTKLDPOCEVEZEHGILGPVRAEVDVIOVRFXNLASRPVSLHAGLSVEKSSGK 1671
 QY 1913 --EDQCGAEPKRFVKNETKTYFKVQVHMAPTKDFCKAWAYFSDVDLXDVHSG 1870
 Db 1672 TYEDDSSEWFKEDNAIQNKTYTWHATRSFGENPGSACRAWAYISAVNPKDIHSG 1731
 QY 1871 IGPLLVCHTNTLPAHQGRQVTVQBFALFTFTFDTKSWYFTENNERNCRAPCNQMDEPT 1930
 Db 1732 IGPLLICRKGTLDEKTNMPVDMREDFVLLFVDFDKKSWYDKPTSRWRASS-----E 1785
 QY 1931 FKENVRFPAINGYNDITPLGLVMAQDQIRWYLLSMGSENIHSHTRSGHFTVRKKEEY 1990
 Db 1786 VKNSHEFPAINGMIYN-LPGLNBYDQEWRLHLLMLGGSRDLHVHFGQVLLNGTOQH 1844
 QY 1991 KVALNLYPGVPEVTEMLPSKAGIWRVECLICEHLHAGMSTLFLVYSNKCOTPLGMASGH 2050
 Db 1845 QLGWVPLLPQGFKTELEKASPGWLLDTEVGEIQACGQTPFLIVDRECKMPGLSTGL 1904
 QY 2051 IRDFOITASQGVQWAPKLARLHYSGSINAW-----STK-EPFSWIKVDLLAPMIHGIK 2104
 Db 1905 IADSQIQASETFWYEPKRLNNGSGSYNANTAEKLTSTFEPFPIQVDMQKEVLLTG 1964
 QY 2105 TQGAQKTSLSIYSFIIMYSLDGKKQWYRGNSGTGLMVFPNGVDSGKIKINFPPII 2164
 Db 1965 TQGAHLYKPYTTFEFCVAYSJDRKNRLEFGNSTRNVNMYFGNSDASTIKENQIDPPV 2024
 QY 2165 ASYIRLHPHYSIRSTRMELMGCDLNSCMPLGMESKASDAQITPASSYFTNMFAT-WS 2223
 Db 2025 ARYIISPTSGYKPKALRELOQCEVNGCSTPLGMESGKIENKQITASSFKKSWGNYWE 2084
 QY 2224 PSKARLHLOGRNAPWQVNNKPEWLOVDPQKTMKVTGTTQGVKSLTSMYVKEPLISS 2283
 Db 2085 PFLARLNAQGRVNAWQAKANNQWLQIDLLKTKITAIVTQCKSLSEMYVKSVTIHY 2144
 QY 2284 SQDGHWTLPFQNGKV--KVFQGNQDSFTPVNSLDPPLLTTRYLRTHPOSWHQIALRME 2341
 Db 2145 SDQGTDKPKYREKSMWVKLFEKNNVNRGHVKFFPPIISRPIRIPKTNQOSIALRUE 2204
 QY 2342 VLGCERQDLY 2351
 Db 2205 LFCC---DMY 2211
 RESULT 7
 CERU HUMAN STANDARD; PRT; 1065 AA.
 AC P0450; Q14063;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
 GN CP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 EN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86259737; PubMed=2873574;
 RA Koschinsky M.I., Funk W.D., van Oost B.A., McGallivray R.T.A.;
 RT "Complete cDNA sequence of human ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5086-5090(1986).
 RN (2)
 RP SEQUENCE OF 1-1006 FROM N.A.

RX MEDLINE=95217183; PubMed=7702601;
 RA Daimon M., Yamatani K., Igarashi M., Fukase N., Kawanami T.,
 RA Kato T., Tomimaga M., Sasaki H.;
 RT "Fine structure of human ceruloplasmin gene";
 RL Biochem. Biophys. Res. Commun. 208:1028-1035(1995).
 RN [3]
 RP SEQUENCE OF 1-40; 549-599; 784-829 AND 919-952 FROM N.A.
 RX MEDLINE=86275241; PubMed=3755405;
 RA Mercer J.F.B., Grimes A.;
 RT "Isolation of a human ceruloplasmin cDNA clone that includes the N-
 terminal leader sequence";
 RL FEBS Lett. 203:185-190(1986).
 RN [4]
 RP SEQUENCE OF 218-1065 FROM N.A.
 RX MEDLINE=86205876; PubMed=3486416;
 RA Yang F., Naylor S.L., Lum J.B., Cutshaw S., McCombs J.L.,
 RA Neberhaus K.H., McGill J.R., Adrian G.S., Moore C.M., Barnett D.R.,
 RA Bowman B.H.;
 RT "Characterization, mapping, and expression of the human ceruloplasmin
 Gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3257-3261(1986).
 RN [5]
 RP SEQUENCE OF 20-1065.
 RX MEDLINE=84113493; PubMed=6582496;
 RA Takahashi N., Ortel T.L., Putnam F.W.;
 RT "Single-chain structure of human ceruloplasmin: the complete amino
 acid sequence of the whole molecule";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
 RN [6]
 RP SEQUENCE OF 158-333; 519-724 AND 858-1065.
 RX MEDLINE=83117800; PubMed=6571985;
 RA Takahashi N., Bauman R.A., Ortel T.L., Dwulet F.E., Wang C.-C.,
 RA Putnam F.W.;
 RT "Internal triplication in the structure of human ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
 RN [7]
 RP SEQUENCE OF 501-905.
 RX MEDLINE=81199407; PubMed=6940148;
 RA Dwulet F.E., Putnam F.W.;
 RT "Complete amino acid sequence of a 50,000-dalton fragment of human
 ceruloplasmin";
 RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
 RN [8]
 RP SEQUENCE OF 907-1065.
 RX MEDLINE=80137543; PubMed=6987229;
 RA Kingston I.B., Kingston B.L., Putnam F.W.;
 RT "Primary structure of a histidine-rich proteolytic fragment of human
 ceruloplasmin. I. Amino acid sequence of the cyanogen bromide
 peptides";
 RL J. Biol. Chem. 255:2873-2895(1980).
 RN [9]
 RP SEQUENCE OF 907-1065.
 RX MEDLINE=80137544; PubMed=6987230;
 RA Kingston I.B., Kingston B.L., Putnam F.W.;
 RT "Primary structure of a histidine-rich proteolytic fragment of human
 ceruloplasmin. II. Amino acid sequence of the tryptic peptides";
 RL J. Biol. Chem. 255:2886-2896(1980).
 RN [10]
 RP SEQUENCE OF 1007-1061 FROM N.A.
 RX MEDLINE=90285218; PubMed=2355023;
 RA Yang F.M., Friedrichs W.E., Cupples R.L., Banifacio M.J.,
 RA Sanford J.A., Horton W.A., Bowman B.H.;
 RT "Human ceruloplasmin. Tissue-specific expression of transcripts
 produced by alternative splicing";
 RL J. Biol. Chem. 265:10780-10785(1990).
 RN [11]
 RP REVIEW
 RX MEDLINE=22049919; PubMed=12055353;
 RA Hellman N.E., Gitlin J.D.;
 RT "Ceruloplasmin metabolism and function";
 RL Annu. Rev. Nutr. 22:439-458(2002).
 RN [12]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).

RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,
RA Lindley P.,
RA "The X-ray structure of human serum ceruloplasmin at 3.1 Å: nature of
RT the copper centres.",
RL J. Biol. Inorg. Chem. 1:15-23 (1996).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers
CC known as type 1 or blue, type 2 or normal, and type 3 or coupled
CC binuclear.
CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -!- DISEASE: Defects in CP are the cause of aceruloplasminemia
CC [MIM:604290]. It is an autosomal recessive disorder of iron
CC metabolism characterized by iron accumulation in the brain as well
CC as visceral organs. Clinical features consist of the triad of
CC retinal degeneration, diabetes mellitus and neurological
CC disturbances.
CC -!- DISEASE: Ceruloplasmin levels are decreased in Wilson's disease,
CC in which copper cannot be incorporated into ceruloplasmin in
CC liver because of defects in the copper-transporting ATPase 2.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13699; AAA51976.1; -;
CC EMBL; D45045; BAA08085.1; -;
CC EMBL; D45044; BAA08084.1; -;
CC EMBL; D45028; BAA08084.1; JOINED.
CC EMBL; D45029; BAA08084.1; JOINED.
CC EMBL; D45030; BAA08084.1; JOINED.
CC EMBL; D45031; BAA08084.1; JOINED.
CC EMBL; D45032; BAA08084.1; JOINED.
CC EMBL; D45033; BAA08084.1; JOINED.
CC EMBL; D45034; BAA08084.1; JOINED.
CC EMBL; D45035; BAA08084.1; JOINED.
CC EMBL; D45036; BAA08084.1; JOINED.
CC EMBL; D45037; BAA08084.1; JOINED.
CC EMBL; D45038; BAA08084.1; JOINED.
CC EMBL; D45039; BAA08084.1; JOINED.
CC EMBL; D45040; BAA08084.1; JOINED.
CC EMBL; D45041; BAA08084.1; JOINED.
CC EMBL; D45042; BAA08084.1; JOINED.
CC EMBL; D45043; BAA08084.1; JOINED.
CC EMBL; D00025; BAA00019.1; -;
CC EMBL; X04135; CAA27752.1; -;
CC EMBL; X04136; CAA27753.1; -;
CC EMBL; X04137; CAA27754.1; -;
CC EMBL; X04138; CAA27755.1; -;
CC EMBL; M13536; AAA51975.1; -;
CC EMBL; J05506; -; NOT ANNOTATED_CDS.
CC PIR; A25443; KUFU.
CC PDB; 1KCM; 12-FEB-97.
CC GlycoSuiteDB; P00450; -;
CC SWISS-2DPAGE; P00450; HUMAN.
CC Sienaa-2DPAGE; P00450; -;
CC Genew; HGNC:2295; CP.
CC MIM; 117700; -;
CC MIM; 604290; -;
CC GO; GO:0005615; Extracellular space; TAS.
CC GO; GO:0004322; Ferroxidase activity; TAS.
CC GO; GO:0006879; P:iron ion homeostasis; TAS.
CC InterPro; IPR001117; Cu-oxidase.
CC InterPro; IPR002355; Cu_ox_copper_BS.

DR InterPro; IPR008972; Cupredoxin.
DR Pfam; PF00394; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER OXIDASE2; 1.
KW Transport; Ion transport; Copper transport; Oxidoreductase; Plasma;
KW Metal-binding; Copper; Repeat; Signal; Glycoprotein; Polymorphism;
KW 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 1065 CERULOPLASMIN.
FT F5/8 TYPE A 1.
FT DOMAIN 20 357 PLASTOCYANIN-LIKE 1.
FT DOMAIN 20 200 PLASTOCYANIN-LIKE 2.
FT DOMAIN 209 357
FT DOMAIN 370 718 F5/8 TYPE A 2.
FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
FT DOMAIN 570 1061 F5/8 TYPE A 3.
FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
FT CARBOHYD 138 135 N-LINKED (GLCNAC. . .).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .).
FT DISULFID 174 200 PROBABLE.
FT DISULFID 276 357 PROBABLE.
FT DISULFID 534 560 PROBABLE.
FT DISULFID 637 718 PROBABLE.
FT DISULFID 874 900
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 180 180 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 182 182 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 994 994 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 997 997 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 999 999 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1039 1039
Query Match 10.6%; Score 1322; DB 1; Length 1065;
Best Local Similarity 19.2%; Pred. No. 1.6e-60;
Matches 398; Conservative 203; Mismatches 417; Indels 1054; Gaps 29;
QY 5 LSTCFPLCLLRFCFATRRYLYGAVELSDVMQSLDGE---LPVDARPPRPVKPFPPNT 61
DB 6 LGIFLFLCSTP-AWAKEKHVYIGIETTWYD-ASDHGKKLISVDEHSHNIYLONGPDRI 63
QY 62 SVVYKTLFVEPTDHLFNIAKPRPPWMLGLGTIOAEVDTVVITKMAHSPVSHVANG 121
DB 64 GRLYKKALYLQYTDFTFRITIEKPVMLGFLGPIIKAEITGDKVYVHLKMLASRPYTHSG 123
QY 122 VSYWKASEGAEYDQTSQREKEDDKVFPGGSHTYVMQVLKENGPMASDPLCLTYSYLSHV 181
DB 124 ITYKSHGAIYPDNTIDFQRAADKVPGEQYTYMLLATEEQSPGEGDGCNVTIYHSHI 183
QY 182 DLVKDLNSGLIGALLVCREGSLAKEKTQTL-HKPIILLFAVDFEGKSWHSE-----TKNS 234
DB 184 DAPKDIASGLIGPLIICKDLSLCKEKHIDREFVVMFVSDENFWSWYLEDNKITYCSEP 243
QY 235 LMDRDRAASARAPKMTVNGYVNSRSLPGIIGCHRSKSVYVHVIOMGTTPVSHSIFLEGT 294
DB 244 EKVDKDNEDFQENRMYSVNGYTFGSLPLGLSMCAEDRKVWLPFGMGNEVDVHAAFFHQA 303
QY 295 FLVNHQAASLEISPIITFLTAQTLMDLQGLLSCHISSHQHDGMEAYVVKVDSCEPBPQL 354
DB 304 LTNKNYRIDTINLFPATLFDAYVAQNPGEWMLSCQNLNHLKAGLQAFPOVQEC----- 357
QY 355 IMKNNEEAEDYDDDLTDSMDVVRFDNDSFPIQIRSVAKKPKTWWHIAAEEDWDY 414
DB 358 ---NKSSSKD-----NIRGKEVRH-----YYIAAEELWNY 385
QY 415 APL-----VLAPDDRYSKSOYLNANGPQIRGRKYKVRFMAYTDTF---KTREAIQ 462
DB 386 APSGIDIFTKENTLAPGSDS--AVFFQGTTRIGSYKKLVRYREYTTDASFNKRERPEE 443
QY 463 HESGILGPLLYGEGVDTLLLIFFKQASRPVNIYPHGI-----TDVRPLY---SRRLP 511

Db 444 EHLGILGEVIAEVDGDIRVTFHFKGAYPLSIPIGVRFKNGEGITYSYNPNQSRSP 503
QY 512 KGVKHLKDFILPGEIEFKYKWTIVVEDGPKSDPRC:TRYYSFVNMERDLASGLIGPLL 571
Db 504 PSASH-----VAPTETTYEWTVPKEVGPINADPVCLAKYYSADVPTKDIFTGLIGPMK 558
QY 572 ICYKESVDQKGNQMSKRNVLFSVFDENRWSYLITENIQFPLNPAGVOLEDPEPQASN 631
Db 559 ICKGSLHANGKQKDVDFEFLPTFVDFENESILLEDNIRNETTAPDQVCKEDDFQESN 618
QY 632 IMHSGVWVFDLSQ-LSVCLHEVAVWYLSIGATDF:SVFPGCYTFKHKMVMYEDTILTF 690
Db 619 KMSNNGFMGNQCLTNCKGDSVVWVLFSGNADVHGIFSGNTYLWRGERDUTANLF 678
QY 691 PFSGETVMSMENFGLWILGCHNSDFNRGWTALLKVSSCDKNTGDYVEDSYEDISAYLL 750
Db 679 POTSILHMFDPTEGTFNVECLTTDHYTGMKQKYTVNQC----- 718
QY 751 SKNAIEPRFSQNSRHPSTKQFNATTIPENDIEKTDPFPAHRTMPKIQNVSSDILL 810
Db 719 ----- 718
QY 811 MLLQSPTPHGLSLSLDQEAKEYTFSDSPGAIDSNNSLSEMTFRPQLHHSQDVVFTP 870
Db 719 --RROSE-----DSTYLGERTY----- 735
QY 871 ESGLOLR-NEKIGTTAAATELKKLDFKVSSTNNLIISTIPSNLAAGTNTSSLGPPSPV 930
Db 736 -----IAAVEVE----- 742
QY 931 HYDSOLDTTLFGKSSPITESGGLPSLSENNDSKLLSGLMNSQSSWGKNVSTESGR 990
Db 743 ----- 742
QY 991 LFKGKRAHPALLTKONALFKVISILKTKNTSNNSATNRKTHIDGPSILLIENSPVMQN 1050
Db 743 -----WD- 744
QY 1051 ILESDETFKVTPLIHDRMLMDKNATALRLNMSNKTTSKNNMVMQCKEPIPDAGN 1110
Db 745 ----- 744
QY 1111 PMSFFKMLFLPESARWIOETHCKNSLSGQSPSPQLVSLGPEKSVQGNFLSEKNV 1170
Db 745 ----- 744
QY 1171 VGKGEFTKDVGLKEMVFPSSRNLFITNLDLHNENHNOEKKIOEBIEKKTILIQENVVL 1230
Db 745 ----- 744
QY 1231 POIHTVTGKFMKNLFLSTRONVEGSDGAYAPVLQDFRSLNDSTNRTKHTAHFSKK 1290
Db 745 -----YSP----- 747
QY 1291 GBEENLEGLGNQTKQVEKYACTRISPNISQGNFVTOQRKALKQFLPLEBETELEKRI 1350
Db 748 -----QR----- 749
QY 1351 IVDDTSTQWSXMKHILTPSLTIQIDYNEKEKGAITQSPISDCLTRSHSIPQANRSPLEPIA 1410
Db 750 -----EWEKELHLJ----- 758
QY 1411 KVSSFPSIRPIYLRVLFDQNSSHLPAASYRKXKDSGVQESSHFLQGNKKNLSIAITLLE 1470
Db 759 -----QEONVSNATL----- 768
QY 1471 MTGQREVGLSGTSATNSVYKKNVTVLKPKDLPTKSGKVELLPKVHIYKOLFPPTETS 1530
Db 769 ----- 768
QY 1531 NGSPGHLDLVEGSLLOCTEGAIKWNEARFKVFFLRVATESSAKTSPSKLPLDPLANDNHY 1590

Db 769 ----- 768
QY 1591 GTOIPXEENKSOEKSPEKTAFAKKKDTILSLNACSNHAIANAINEGQNKPEIETWAKQGR 1650
Db 769 ----- 768
QY 1651 TERLCSQNPVLKRHOREITRITLQSQDEIDYDDTISVEMKKEDFDIYDEDENQSPRSF 1710
Db 769 -----DKGZ----- 772
QY 1711 OKTRHYFIAAVERLDWYGMSSPHVLNRAQSGSVPEQKKVVFQBTDCSFTQPLRYGE 1770
Db 773 -----FYIGS-----KIKKVVIRQYIDSTFRVPVERKA 800
QY 1771 LNEHLGILLGIVIRAEVEDNIMVTPRNOASRPYGFYSSLSIYEEDQRCQCAPRKNFVK--- 1827
Db 801 EEEHLGILGQLHADVGDKVLIKFNWATRPYSIHA-----HGQVETESTVTPTL 850
QY 1828 PNETKTYFWKVQHMAFTKDBEFCCKAWAYFSDVLEKDVHSGSLIGPLLVCHTNTLNAHG 1887
Db 851 PGEILTYVWKIPERSGAGTEDSACIPWAYYSTVDQVLDLYSLGSLGGLIVCRPRPYLKVFPN 910
QY 1888 RQVTVQBPALFFTFIDETKSNYFTENMERNCRAPCNIQMEDPTFKNYRPHAIYIMDT 1947
Db 911 RRKL--EPALLFLYFDENESWYLDNITKYSDDHEKVKKDDBEFIESKMAINGRMPGN 968
QY 1948 LPGLVMAQDQIRWYLLSMGSENHSHFSGHVFTVRKKEEYKVALYNLYPGVPETVEM 2007
Db 969 LOGITMVGDEVNWLKMGNEIDLHTVHFHGHFQYKRGVYSSDVFDPFGTYQTLEM 1028
QY 2008 LPSKAGIWRVECLIGELHAGNSTLFLVYSNK 2039
Db 1029 FPRTEGILLCHVTDHAGMETTYTVLQNE 1060

RESULT 8
CERU RAT ID CERU RAT STANDARD: PRT: 1059 AA.
AC P13635: Q64719:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Rattus norvegicus (Rat).
OC Sukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Sutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver, and Lung;
RX MEDLINE=90237081; PubMed=2332446;
RA Fleming R.E., Gitlin J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
tissue-specific gene expression during development.";
RL J. Biol. Chem. 265:7701-7707(1990).
RN [2]
RP SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.
RX TISSUE=Liver;
RC MEDLINE=87137545; PubMed=3818625;
RA Aldred A.R., Grimes A., Schreiber G., Mercer J.P.B.;
RT "rat ceruloplasmin. Molecular cloning and gene expression in liver,
choroid plexus, yolk sac, placenta, and testis.";
RL J. Biol. Chem. 262:2875-2878(1987).
CC -I- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
molecule) glycoprotein found in plasma. Four possible functions
are ferroxidase activity, amine oxidase activity, copper transport
and homeostasis, and superoxide dismutase activity.
CC -I- FUNCTION: May also play a role in fetal lung development or
pulmonary antioxidant defense.
CC -I- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -I- COFACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
the multicopper oxidases which contain three distinct Cu centers


```
QY 1057 EFKKVTELTIDRMLMDKATLRNLNMSNKTTSKONKEMVQKKSGEIPDPDAQPMDSFF 1116
Db 733 ----- 732
QY 1117 KMLFLPESAWIORTGCKNSLNSQGSPKQLVSLGPEKSGVEGQFLSEKNKVVVGKGEF 1176
Db 733 ----- 732
QY 1177 TKDVLKEMVPPSRNLELTNLDNLHENHNNHNOBKJQEBIEKKEKTLIQENVVLPOIHTV 1236
Db 733 ----- 732
QY 1237 TGTNFKMLFLSTRQNVESYDCAVAPVLQDFRSLNDSNRRKXKTAHFSKKGBEENL 1296
Db 733 ----- 741
QY 1297 EGLGNQTKQIVKXVACTFRISNTSQQNFVQRSKRLKQFRLPLETELEKRIIVDDTS 1356
Db 742 ----- 742
QY 1357 TQWKNMKHLFTLTQIDYNEKEKGAITQGPLSDCLTRSHSIPQANRSPDPIAKVSFP 1416
Db 743 RDWEMELHEL----- 752
QY 1417 SIRPIYLRVLFDQNSHLPAASVKKDSGVQESSHFLQAKKNLSLAITLEMTGQDR 1476
Db 753 ----- 762
QY 1477 EVGSGTSAATNSVTKYKVENTVLPKPLPKTSKGVLLPKVHIYQKDLFPPTETSGSPGH 1536
Db 763 ----- 762
QY 1537 LSLVEGSLQGTGAIKMNEANRPGKPLRVATSSAKTPSKLLDPLANDHNYGTQPK 1596
Db 763 ----- 762
QY 1597 EEWKSEKSPKTAFKKXDTILSNACSHALAINQGNKPEIBVTWAKQGTBLRCS 1656
Db 763 ----- 762
QY 1657 QNPVLKRQREITRTTLOSOREIDYDDTISVEMKKEDFDIYDEENOSPRSFQKTRH 1716
Db 763 ----- 766
QY 1717 YFTAAVERLDYGMSSPHVLNRPAOSGVPQKVVQFQFTDGTSTQPLNROGLNEHLG 1776
Db 767 FFIGS----- 800
QY 1777 LLGPYIRAEVDNIMVTERNOASRPYSFSSLSIYBEDQCGAEPRKPFVK---PNETKT 1833
Db 801 MLGFLIHADVGAKVVKVFKMARTFYSIA-----HGVTKSSTVAPTLPGVRT 850
QY 1834 YFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLVC---HTWTLNPAHQV 1890
Db 851 YIQIPERSGAGTSDSPICWAYISTVDRVKDLYSLGLIGLIVCRKSVKVFNPK----- 905
QY 1891 TVQEFALFTTIFBTKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHAINYMDPLG 1950
Db 906 KQMEFSLFLVFDENESWYLDNNTNTPDHPKDNKDNBEFIESKQKHANGKMGFNLQ 965
QY 1951 LNPACQDQIRKYLKSMGNSNIHSIFSGHVFTVRKXKEVKMALNLYPGVFETVELMPS 2010
Db 966 LTHVGVENMYVWANGNEIDLEIVHFGHSFQYKRGHSDDVDFPFGTYQIEMFPQ 1025
QY 2011 KAGIWRVECLIGEHLHAGMSTFLVYSNK 2039
Db 1026 TPGTWLHCHVTDHAGMSTVTVLVNQ 1054
PRT: 1062 AA.
ID CERU_MOUSE STANDARD;
RESULT 9
CERU_MOUSE
ID CERU_MOUSE
```

```
AC Q61147;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Klomp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE SPECIFICITY.
RA MEDLINE=96294736; PubMed=9690795;
RA Klomp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;
RT "Ceruloplasmin gene expression in the murine central nervous system.";
J. Clin. Invest. 98:207-215(1996).
CC -!- FUNCTION: Ceruloplasmin is a blue, copper-binding (6-7 atoms per
CC molecule) glycoprotein found in plasma. Four possible functions
CC are ferroxidase activity, amine oxidase activity, copper transport
CC and homeostasis, and superoxide dismutase activity.
CC -!- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -!- COPACTOR: Binds 6 Cu-ions per molecule. This protein belongs to
CC the multicopper oxidases which contain three distinct Cu centers
CC known as type 1 or blue, type 2 or normal, and type 3 or coupled
CC binuclear.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues, including liver,
CC eye and brain.
CC -!- SIMILARITY: Contains 3 F5/8 type A domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49430; AAB07996.1; -.
DR HSSP; P00450; IKCW.
DR MGD; MGI:98476; Cp.
DR InterPro; IPR001117; Cu-Oxidase.
DR InterPro; IPR002355; Cu_ox_copper_BS.
DR InterPro; IPR008972; Cuprofoxin.
DR Pfam; PF00394; Cu-Oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
DR Transprot; Ion transport; Copper transport; Oxidoreductase; Plasma;
KW Metal-binding; Copper; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 1062
FT F5/8 TYPE A 1.
FT DOMAIN 20 356
FT PLASTOCYANIN-LIKE 1.
FT DOMAIN 20 199
FT PLASTOCYANIN-LIKE 2.
FT DOMAIN 208 356
FT F5/8 TYPE A 2.
FT DOMAIN 369 713
FT PLASTOCYANIN-LIKE 3.
FT DOMAIN 369 555
FT PLASTOCYANIN-LIKE 4.
FT DOMAIN 565 713
FT F5/8 TYPE A 3.
FT DOMAIN 725 1057
FT PLASTOCYANIN-LIKE 5.
FT DOMAIN 725 896
FT PLASTOCYANIN-LIKE 6.
FT DOMAIN 904 1057
FT BY SIMILARITY.
FT DISULFID 173 199
FT BY SIMILARITY.
FT DISULFID 275 356
FT BY SIMILARITY.
FT DISULFID 529 555
FT BY SIMILARITY.
FT DISULFID 632 713
FT BY SIMILARITY.
FT DISULFID 870 966
FT COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 120 122
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 122 122
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179
FT COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181
FT COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 990 990
```

```

FT METAL 993 993 COPPER (TYPE 2) (BY SIMILARITY)
FT METAL 995 995 COPPER (TYPE 3) (BY SIMILARITY)
FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY)
FT METAL 1036 1036 COPPER (TYPE 1) (BY SIMILARITY)
FT METAL 1037 1037 COPPER (TYPE 3) (BY SIMILARITY)
FT METAL 1041 1041 COPPER (TYPE 1) (BY SIMILARITY)
FT METAL 1046 1046 COPPER (TYPE 1) (BY SIMILARITY)
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 226 226 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 583 583 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 625 625 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 757 757 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 922 922 N-LINKED (GLCNAC. .) (POTENTIAL)
SQ SEQUENCE 1062 AA; 121159 MW; F3F52ED09A239F16 CRC64;

Query Match 10.0%; Score 1243; DB 1; Length 1062;
Best Local Similarity 18.9%; Pred. No. 1.9e-56;
Matches 393; Conservative 196; Mismatches 432; Indels 1054; Gaps 30;

QY 10 FLCLLRFCF-----SATRYILCAVELSWDMQ--SDLGELPVDARFPRVPKSPFFN 60
DQ 3 FLLSTFIFLYSSALARDKHYFIFIGITEAVWDYASGTEKKLIIISVDTSQSNFYLONGDR 62
QY 61 TSUVYKTLFVEFTDHLFNIAKPRPPWGLLGPITQAEVYDTVTITLKNMASHPVSLHAV 120
DQ 63 IGRKYKALFYDTGTSTKIDKEAMGLFGLPVIAEVEDKVYVHLKNLASRIYTFHAR 122
QY 121 GVSYKASGAEYDQTSOREKEDDKVPFGSGHTYVWQVLKENGPMADPLCLTYSLSH 180
DQ 123 GVTYKEYEGAVPDNTDFQADDKVLPGQGVVYVTLHA-NEPSPGEGDSNCVTRIYHSH 181
QY 181 VDLVKDLSGLICALLVCAGSLAKETQL-HKRIILFAVDEGSKHSETKNSLM--- 236
DQ 182 VDAPEDIAAGLGLPLCKKGLYKEKEKNIDQEFVLMFVSVDENLSWLENIITFCSE 241
QY 237 ---QDRDAASARAWPMHWNGVNRSLPGLICGRKSVYVHWIGMTTPEVHSIFLEGH 293
DQ 242 PEKVDKNEDFQESNMYSYNGTYFGLSLGNCADRVKWLFGMGNEVDVHSAPFHQ 301
QY 294 TFLVNRHQASLEISITPLTQTLMDLQGLLSCHISHOHQMGAEVAVKVDSPPEEPQ 353
DQ 302 ALTSRYNQTDIINLPATLIDAVVAQNEGVWMLSCQNLNHLKAGLQAFQVQDC----- 356
QY 354 LINKNBERAEDDDDLTDSMDVVRDDDNSPSFIQIRSVAKGHPKTWVHYTAAREEDWD 413
DQ 357 ---NKPSSKD-----NIRGKVRH-----YYIAAEVAVXN 383
QY 414 YAP--LVLAPDDR-----SYKSYLNNGFQIRGRKYKVRFMAYTDTF---KTREAIQH 463
DQ 384 YAPSGDIFTEEBKLTASGSDSGVFFEQGATRIGSYKKMAYREYTDGSPFNKRGPDDE 443
QY 464 ESGILGPLYGEVGDITLLIIFKQASRPYNIYHGIT-----DVRPLYSRKLPKGVKHXD 519
DQ 444 HLGILGPIVIAEYVGDITKVTFHNKGQHUSIQMGVSFTAEENGTYG---PPGASSQQA 500
QY 520 FP-ILPGEIIFYKWTIVDEGPKSDPRCLTRYVSSFWANERDLASGLIPLLCYKESV 578
DQ 501 ASHVAPEKTYTWTVPKNGPTVADVC-SKMYISAVDPTKOLFGLIGPMKICKGSL 560
QY 579 DQGNQIMSKRNVIILSVFDENSRWYLTENIQRFENPAGVQLEDEPFOASNMHSING 638
DQ 561 LADGRQXVDKEFFLPTVFDENESILLDDNIRMTHTAPQVDKEDBDPQESNKMHSNG 620
QY 639 YVFDLSQI-SVCLHEVAYWILSICATDFLSVFFSGYTKHKMYVEDTLTLPPSGSETV 697
DQ 621 FMYGNQSWPHMCLGESITVWYLFSAENADVHGIFYSGNTYLCRGEEDTANLPHKSLTL 680
QY 698 FMSMENPGLWILGCHNSDFNRNGMTALLKVSXCDKNTGDDYEDSYEDISAKLISKNAIE 757
DQ 681 LMPNEDTKGTFDVECLTIDHYTGMKQKXTVNOQOR-----QFEPFTVYL----- 724
QY 758 PRSPQNSRRPSTRQKQFNATTIPENDIEKTDPDFAHRTMPKIQNVSSSDLLMLLRQSP 817

```

```

725 ----- 724
QY 818 TPHGLSLSDIQAQAK(EFFSDDDPSGGAIDSNNSLSEMTHERPQLHSGDMVFTPEGLQLR 877
725 ----- 724
QY 878 LNEKLGTTAAATELKCLDFKVSSSTNNLIISTIPSDNLAAGTNTSSLGPPSPMPHYDSOLD 937
725 ----- 724
QY 938 TTLPGKKSSPLTBSGGPLSLSEBNNDSKLLESGLMNSQESSGKNGKVSSTESGRLFKGRKA 997
725 -----GERT 728
QY 998 HGPALLTKDNALFKVSTISLLKTKNTKNSATNRKXTHIDGSPSLIENSFVWQNLIESDTE 1057
729 Y-----YVDA----- 733
QY 1058 FKVYPLIHDRMLMDKNATALRLNMSNKTTSKKNMEMVQOKKEGPDPDAQNDPMSEFFK 1117
734 ----- 733
QY 1118 MLFPEGARWIORTHGKNSLNSGOGSPKQIIVSLGPEKSVEGQNFLEBKKNVVGKGBFT 1177
734 ----- 733
QY 1178 KDVGLKEMVPPSRNLFILTNLDNLHENNTHNQERKIOBEIEKKETLTOENVLPLQIHTVT 1237
734 ----- 733
QY 1238 GTRKFMKNLFLSTQNVGSDGAYAPVLQDFRSLNDSTNRKHTAHFASKGEEENLE 1297
734 -----VEVEWD--YS- 742
QY 1298 GLGNCTKQIVKXYACTTILSPNTSQNFVQTSKEALKQPLPLEETELEKRIIVDDTST 1357
743 -----SR 744
QY 1358 QWSKNMKHLTFTLTQIDYNEKEGAIQTSPSLDCLTRSHSIPQANRSPPLPIAKVSSPFS 1417
745 AWEKELHHL----- 753
QY 1418 IRPIYLTRVLFODNSHLPAASRYKKDSGVQBSSEHFLQAKKNNLSAILILEMTGDQRE 1477
754 -----QEQNYS----- 759
QY 1478 VGLSGLTSATNSVTYKXVENTVLPKDPLPKTCKVBLPKVHIYQKDLFPTETSGSPGHL 1537
760 ----- 759
QY 1538 DLVEGSLLOGTGGAIKWNEANRPGKVPFLRVATSSAKTPSKLLDPLAWDNHYGTQIPKE 1597
760 ----- 759
QY 1598 EWSQEKSPKTAFFKKDTILSNACESHAAINEGONKPEIEVWTWAKQGRTERLCSQ 1657
760 ----- 759
QY 1658 NPPVLKXHQREITRTTLQSDQBEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKTRHY 1717
760 -----NVFLDKXERFI----- 770
QY 1718 FIAAVERLMDYGMSSSPHVLNRQAQSGVQPKVVFQBFQSGFTQPL-YRGELNEHLG 1776
771 -----GS--KYKVVYRQDTSSSFEQVKRAEEDHLG 802
QY 1777 LLOPYTRAEVEDNIMVTFRQAASRPYSFYSSLSIYEDQROGAEPKNTF---KPNETKT 1833
803 ILGPIIFANVGDVKVVPFKNMATRPYSIHA-----HGVTESSTVPTLPOEVAT 852
1834 YFWKVQHHWAPTDEDFCKAWAYFSDVLEKQVHSGLIGPLLYC---HTNLTLMFAHRQV 1890
853 YTWQIPERSGAGREDSACIFWAIYSTVDRVKDLYSLIGPLIVCRKSYVKVPSPK----- 907

```


QY 1891 TVQFALEFFITFDTKSYFTENMERNCRAPNTQMSDPTPKENYRPHAINGYIMDTLP 1950
Db 908 KXMEFFLLFVFDENESYLDNKTYSERKYNKONEEFLESNKMEALNGKFGNLOG 967
QY 1951 LVMAQDQIRWYLLSMGSENEHSHSGHFTVRKKEEKWALYNLPVGFVEMLPS 2010
Db 968 LTMVQKQEVNWMGMGNEDLHTVHFHGFQYKRGVYSDVDFLPGTYQILEMFPQ 1027
QY 2011 KAGLWVSECLGERHAGMSITFLVYSNKCQTPLG 2045
Db 1028 TPGIWLHCHVTDVHAGMATYIVLPVEQETKSG 1062
RESULT 10
MFGM_RAT STANDARD; PRT; 427 AA.
ID MFGM_RAT
AC P70490;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
acetyl GD3 ganglioside synthase) (AGS) (MFGM).
GN MFG38 OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96374422; PubMed=8780713;
RA Ogura X., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RT "Cloning and expression of cDNA for O-acetylation of GD3
ganglioside."
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -!- FUNCTION: May be involved in phospholipid binding. Seems to
participate in the O-acetylation of GD3 ganglioside sialic acid.
CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein.
CC -!- TISSUE SPECIFICITY: Spleen, lung, heart, brain and muscle.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation.
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D84068; BAA12210.1; -.
CC PIR; JCA915; JCA915.
CC HSP; P00740; IEDM.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR000421; FA5C.
CC InterPro; IPR008979; GalBind like.
CC InterPro; IPR006210; IEGF.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC SMART; SMO0181; EGF_2.
CC SMART; SMO0231; FA5C_2.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS00026; EGF_3; 2.
CC PROSITE; PS01285; FA5C_1; 2.
CC PROSITE; PS01286; FA5C_2; 2.
CC PROSITE; PS00022; FA5C_3; 2.
CC PROSITE; PS00022; FA5C_3; 2.
KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 427 LACTADHERIN.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.
FT

FT DOMAIN 111 267 F5/8 TYPE C 1.
FT DOMAIN 272 427 F5/8 TYPE C 2.
FT DISULFID 28 39 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 111 267 BY SIMILARITY.
FT DISULFID 254 258 BY SIMILARITY.
FT DISULFID 272 427 BY SIMILARITY.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 427 AA; 47413 MW; EA8C8631F3E36047 CRC64;
Query Match 5.3%; Score 663; DB 1; Length 427;
Best Local Similarity 40.3%; Pred. No. 4e-27;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;
QY 2018 ECLIGEHUAGKSTLFLVYSNK-----CQPLGWSAGHIRDFOITASGQY-- 2062
Db 78 KCLVETDQRG--DIFTEYICQCPVYSGIHCELCGCTKLGEGGAIADSQISASVYMG 135
QY 2063 ----GOWAPKLARLHYSGSINAW--STKEPFSWIKVDLLAPMIHGKIQTQAGRKFSLY 2116
Db 136 FMLQRPQPELRLVYTGIVNANTASSYDSKPWIOVDFLRKVRVSGVMTQGASRAGRAY 195
QY 2117 ISQFIIMYSLGKKWQTYRGNSGTGLMVFFGNVDSGGIKHNFNPITARYILRHETHYS 2176
Db 196 LKTEKVAVSLDGRPEFTQDESSTGDKFEMGMQDNNSLKINMENPTLEAQYIRLVPVSC 255
QY 2177 IASTLRMELMGCLNSCSMELGMSKASDAOITASSYFT--NMFA-TWSPSKARLHLOG 2233
Db 256 RCTLRFLFELGELGCGEPLGLKNTIPDSQITASSSKYTNLRAFQWPHGLRLDNOG 315
QY 2234 RSNVRPQVNNPEKLQVDFQKTMKVTGTTQGVKSLITSMYKFLISSODGHWTLF 2293
Db 316 KINAWTAQNSAKEMLOVDLQKVKVTGIITQGDGFHIGYVASYKVAHSDGQWTVY 375
QY 2294 FQNGKVKVFGQNDQSFVFNLSLDPPLRLRYLRHPQSWHQAIALMEVLGC 2345
Db 376 EEGTSSKVPQGNLDNNSHKNFEEKPFPARYVRVLPVLSWHNRITLLELLGC 427
RESULT 11
MFGM_MOUSE STANDARD; PRT; 463 AA.
ID MFGM_MOUSE
AC P21956; P97800;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM).
GN MFG8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC TISSUE=Mammary gland;
RX MEDLINE=91046006; PubMed=2122462;
RA Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RT Srinivasan U., Parry G.;
RT "cDNA cloning of a mouse mammary epithelial cell surface protein
reveals the existence of epidermal growth factor-like domains linked
to factor VIII-like sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN [2]
RP SEQUENCE OF 23-463 FROM N.A.


```
DR Pfam; PF00754; F5_P8_type_C; 2.
DR SMART; SM00181; EGF; 2.
DR RA SMART; SM00231; FAS8C; 2.
DR DR PROSITE; PS00022; EGF_1; 2.
DR DR PROSITE; PS01186; EGF_2; 2.
DR DR PROSITE; PS00026; EGF_3; 2.
DR DR PROSITE; PS01285; FAS8C_1; 2.
DR DR PROSITE; PS01286; FAS8C_2; 2.
DR DR PROSITE; PS00022; FAS8C_3; 2.
KW Glycoprotein; Repeat; EGF-like domain.
FT DOMAIN 2 41 EGF-LIKE 1.
FT DOMAIN 44 88 EGF-LIKE 2.
FT DOMAIN 91 247 F5/8 TYPE C 1.
FT DOMAIN 252 409 F5/8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 409 AA; 45725 MW; BOC07AF80029927A CRC64;

Query Match 5.2%; Score 650; DB 1; Length 409;
Best Local Similarity 39.3%; Pred. No. 1.8e-26;
Matches 139; Conservative 62; Mismatches 125; Indels 28; Gaps 6;

QY 2018 ECLIGEHLHAGMSLFLVYSNK-----CQPLGMSAGHIRDFOITASGOY-- 2062
Db 58 ECEVIDDAHRS--DVFEYICKPHGYTGHCILCNAPLGMETGATDFQISASSHLG 115
QY 2063 ---GOWAPKLARIFYGGSINAW--STKPSPSWKVLLAPMIHGHKTQARQKFSLY 2116
Db 116 FMGLQRAWPELRLHRAQIVNAWTASNDYRNPQVNLRLRRVTVGVVGTQASAGSAEY 175
QY 2117 ISQFIHYSLDCKKQWTRGNSGTGLVFFGNDVSSGKHNIENPPIIARVIRLHPHYS 2176
Db 176 MKTFKVAISTDGRKFQFIQGAESGDKIFMGLNDSGLKVLNLFVPELVQVRLVPIICH 235
QY 2177 IRSTLRMLMGCDLNSCMLPMSKATSDAQITASSYFTN---MFAWSPSKARLHLQG 2233
Db 236 RCTILRPFLGCLSGCAEPLGLKNDTTPNKGITASSFYRTWGLSFAWSPFYARLDNQ 295
QY 2234 RSNARPPVNNKPELVDFQKTKMTGVTGQVKSLTSTMYKVELTSSSQDGHQWTLF 2293
Db 296 KFNWTAQSNASASEWLQIDLSQSRVGTGITQARDPQHIQVAAKYAVYSDDGVSWTET 355
QY 2294 PQNGKV--KVFQGNDSFTPVNSHDPPLLDYLRHPSQVHQAIRMEVLGC 2345
Db 356 RDQGALEKIFPNLDNNSHKQKMFETPLFRFVRLPVANHRITLRLVELLGC 409

RESULT 13
ID MFGM_BOVIN STANDARD; PRT; 427 AA.
AC Q95114; F79344; Q27959;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
DE (MG57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
DE SP47) (BP47) (Components 15/16).
GN MFG88.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
ON NCBI_Taxid=9913;
OX [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RC STRAIN=Holstein; TISSUE=Mammary gland;
```

```
RX MEDLINE=97008954; PubMed=8856064;
RA Hvarregaard J., Andersen M.H., Berglurd L., Rasmussen J.T.,
RA Petersen T.E.;
RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
RT milk fat globules.";
RL Eur. J. Biochem. 240:628-636(1996).
RN [2]
RP SEQUENCE OF 18-427 FROM N.A.
RC TISSUE=Mammary gland;
RX MEDLINE=96125736; PubMed=8541316;
RA Aoki N., Kishi M., Taniuchi Y., Adachi T., Nakamura R.,
RA Matsuda T.;
RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
RT monoclonal antibodies raised against bovine milk fat globule
RT membrane.";
RL Biochim. Biophys. Acta 1245:385-391(1995).
RN [3]
RP SEQUENCE OF 19-427 FROM N.A.
RC TISSUE=Testis;
RA Esslin M.A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
RC TISSUE=Milk;
RX MEDLINE=93250576; PubMed=8485470;
RA Mather I.H., Banghart L.R., Lane W.S.;
RT "The major fat-globule membrane proteins, bovine components 15/16 and
RT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
RT containing epidermal growth factor-like and factor V/VIII-like
RT sequences.";
RL Biochem. Mol. Biol. Int. 29:545-554(1993).
CC -!- FUNCTION: Probably associates with phospholipids on the surface of
CC mammary epithelial cells and milk fat globules. Zona pellucida-
CC binding protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q95114-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q95114-2; Sequence=VSP_001398;
CC -!- TISSUE SPECIFICITY: Milk and spermatozoan.
CC -!- PTM: THE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```
DR InterPro; IPR006210; EGF.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFBL00D.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00026; EGF_3; 1.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00022; FAS8C_3; 2.
KW Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT SIGNAL 1 23
FT CHAIN 24 387
FT CHAIN 202 397
FT CHAIN 268 317
FT CHAIN 24 67
FT DOMAIN 70 225
FT DOMAIN 230 387
FT SITE 46 48
FT SITE 27 38
FT DISULFID 32 55
FT DISULFID 57 66
FT DISULFID 70 225
FT DISULFID 212 216
FT DISULFID 230 387
FT CARBOHYD 238 238
FT CARBOHYD 325 325
FT CARBOHYD 329 329
FT CARBOHYD 350 350
FT CARBOHYD 387 387
FT SEQUENCE 387 AA; 43123 MW; 286571DEC83782D CRC64;
Query Match 4.7%; Score 588; DB 1; Length 387;
Best Local Similarity 37.3%; Pred. No. 2.6e-23;
Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;
QY 2006 EMLPKAGIWRVCELGHEHAGMSTFLVYSNKOTPLGMASGHIRDFQIFASG---- 2060
DB 48 DFFPS-----YTCTLKG---YAGNHC-----ETKVEPLMGNGNLIANSQINASSVRVTF 95
QY 2061 -QYGOWAPKARLHYSGSINAW--STKEPFSIKVDLLAPMIHGIKTQGARQKFSSTLY 2117
DB 96 LCLQWVPELRLNRCAGVNAWTPSSNDNPMIQVNLRLRYKVVTVGTQASRLASHEYL 155
QY 2118 SQFILYLDGKKQWQYRNGSTGTLMVPPGVNDSGKIKNINPPIIARYILRHPTHSI 2177
DB 156 KAFKVAVSLNCHFEFD-FIHDVNNKKHKEFVGNWNNQNAVHVNLPETPVEAQYVLYPTSCHT 214
QY 2178 RSTLMEIAGCDLNSCMPLGMESKAISDAQITASSYF---TNMPATWSPSKARLHLQ 2233
DB 215 ACTLRFELGCEIAGCANPLGLKNNISDPKQITASSYKTVGLHLP-SWNPFSYARLDKQ 273
QY 2234 RSNARVPQNNKREWLOVDFQKTKVGTGTQGVKSLTSMVYKFLISSODGHQWTLF 2293
DB 274 NFNWVAGSYGNDQLQVLDLSSKEVTGILITQARNFGSVQFVASKVAYNSDANWTEY 333
QY 2294 F--QNGKVKVQGNQDSFTFVNSLDPDLTLRLRHPSQWHQIALRMEVLGC 2345
DB 334 QDPRTGSKIFFGNWNBHSHKGNLPTPIIARYVRLPVAWENRIARLELLGC 387
RESULT 15
ID NRP2_HUMAN STANDARD; PRT; 931 AA.
AC O60462; O14820; O14821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE receptor 2).
GN NRP2 OR VEGF165R2.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX MEDLINE=97470888; PubMed=9331348;
RA Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT "Neuropilin-2, a novel member of the neuropilin family, is a high
RT affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT III."
RL Neuron 19:547-559(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A22).
RX TISSUE=Breast;
RA Soker S., Takashima S., Miao H.-O., Neufeld G., Klagsbrun M.;
RT "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT isoform-specific receptor for vascular endothelial growth factor."
RL Cell 92:735-745(1998).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20309748; PubMed=10748121;
RA Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT form of vascular endothelial growth factor (VEGF) and of placenta
RT growth factor-2, but only neuropilin-2 functions as a receptor for
RT the 145-amino acid form of VEGF."
RL J. Biol. Chem. 275:18040-18045(2000).
CC -!- FUNCTION: High affinity receptor for semaphorins 3C, 3F, VEGF-165
CC and VEGF-145 isoforms of VEGF, and the PLGF-2 isoform of PGF.
CC -!- SUBUNIT: Neuropilin-2 probably forms a heteromeric complex with
CC neuropilin-1 in order to be a functional semaphorin 3C receptor.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=A22;
CC IsoId=060462-1; Sequence=Displayed;
CC Name=A0;
CC IsoId=060462-2; Sequence=VSP_004342;
CC Name=A17;
CC IsoId=060462-3; Sequence=VSP_004341;
CC -!- SIMILARITY: Belongs to the neuropilin family.
CC -!- SIMILARITY: Contains 2 CUB domains.
CC -!- SIMILARITY: Contains 2 F5/8 type C domains.
CC -!- SIMILARITY: Contains 1 MAM domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF022859; AAC51788.1; -
CC EMBL; AF022860; AAC51789.1; -
CC EMBL; AF016098; AAC12922.1; -
CC HSPB; P12259; 1C2T.
CC Genew; HGNC:8005; NRP2.
CC NIM; 602070; -
CC GO; GO:0005624; C:membrane fraction; TAS.
CC GO; GO:0004872; F:receptor activity; TAS.
CC GO; GO:0005021; F:vascular endothelial growth factor receptor. . . ; TAS.
CC GO; GO:0007411; P:axon guidance; TAS.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR000421; FAS8_C.
CC InterPro; IPR008979; Gal_bind-like.
CC InterPro; IPR000938; MAM_domain.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00754; F5_F8_type_C; 2.
CC Pfam; PF00629; MAM; 1.
CC PRINTS; PR00020; MAMDOMAIN.
CC SMART; SM00042; CUB; 2.
```

QY 2257 MKVTGVTQ-----VKSLTSMYKKEFLISSQDGHQWTLFQNGKV---KVFQGNQD 2307
Db 496 KTVKGVIQQARGGDSITAVEARAFVKKFKVSYSLNGKWE-YIQDPTQOPKLFEGNMH 554
QY 2308 SFTPVNSLDPPLTLRYLRHPOSVWHQ-IALRMEVLGCEAQD 2349
Db 555 YDTPDIRFD-PIPAQXRVYPERWSPAGIGMRLEVLGCDWTD 596

Search completed: April 13, 2004, 14:05:28
Job time : 38 secs

DR SMART; SMO0231; FA58C; 2.
DR SMART; SMO137; MAM; 1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01285; FA58C1; 2.
DR PROSITE; PS01286; FA58C2; 2.
DR PROSITE; PS01286; FA58C3; 2.
DR PROSITE; PS00022; FA58C3; 2.
DR PROSITE; PS50060; MAM; 2.
KW Transmembrane, Glycoprotein; Neurone; Signal; Repeat; Receptor;
Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 931
FT DOMAIN 21 864
FT TRANSMEM 865 889
FT DOMAIN 890 931
FT DOMAIN 28 142
FT DOMAIN 149 267
FT DOMAIN 277 427
FT DOMAIN 434 592
FT DOMAIN 642 802
FT DOMAIN 671 674
FT DISULFID 28 55
FT DISULFID 83 105
FT DISULFID 149 175
FT DISULFID 208 230
FT DISULFID 277 427
FT DISULFID 434 592
FT CARBOHYD 152 152
FT CARBOHYD 157 157
FT CARBOHYD 629 629
FT CARBOHYD 839 839
FT VARSPLIC 809 813
FT VARSPLIC 809 830
FT CONFLICT 602 602
FT SEQUENCE 931 AA; 104830 MW; 270CBAB69A0A797C CRC64;
Query Match 3.8%; Score 469.5; DB 1; Length 931;
Best Local Similarity 28.8%; Pred No. 1.3e-16;
Matches 166; Conservative 87; Mismatches 195; Indels 145; Gaps 32;
QY 1847 DEFDC-AWAYFSDVLEK----DVHSLGILGLLVCHNTYL-----NPAHGRQVTVQBEFA 1896
Db 79 EKHCKVDFEIRDGSEADLLGKHCNGIAPPTIISSGSMYIKFTSDYARQCA--GFS 136
QY 1897 LFTTIDETKSWFTENMERNCRAPCIQMEDTEKENYRHALNGYIMDTLRLYMAQD 1956
Db 137 LRYEIR-ITGS-----EDCSKRFSP-KGTIESGFPPEXP----- 170
QY 1957 QRIRWLLSGMSENTHSIHFGSHVFWKKEEYKALYKLYPGVFTVEMLPKAG--- 2013
Db 171 -----HNLDCTFIIAKFKWEILQFL---IFD-LSHDPLQVGE 207
QY 2014 -----IW-----RVECLIGEH-----LHAGMST-----LFLY 2035
Db 208 CKYDWDIDWDGPHVGLPGICKYCTKTPSELRSSTGLSTFTHTDMAVAKGFGSARYLV 267
QY 2036 YSN-----KCTPLGMASGHIRDQITASGOY--GOWAPKLRLHYSGSINAW-----ST 2083
Db 268 HQELENFQCNVPGMESGRIANEQISASTYSQGWTPQOSRLH--GDDNGWTFNLDN 325
QY 2084 KEPTSWIKVDLLAPMIIHGKITQGA--RQKPSLIYSQFIIVYSLDGKKWQTYRGNSTGT 2141
Db 326 K3---YLOVDLRFULTMTAIATQGAISRETQNGYVYKYLEVSTINGEDMWYRHGKNH- 381
QY 2142 LMFVFGNVDSGKIKHINFPPIIARYELHPTHYSIRSTRMELMGCDLNS--CSMPFGM 2199
Db 382 -KVFOANNDATEVVLNKLHAPLLTRFVRIRPQTWHSGIALRLFLPGCRVTDAPCSNMLGM 440
QY 2200 ESKAISDAQITASSYFTNNMTATWSPSKARLHLCQGSNAN---RPQVNNPKEMLOVDFOKT 2256
Db 441 LSGLIADSQISASS--TQFY-LWSPSAARL-VSSRS-GWFPRIPOAQPGEEWLQVDLGT 495

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:27 ; Search time 68.5 Seconds
(without alignments)
10828.953 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415
Sequence: 1 MQLSTCFCLLRPFSA.....WVHQIALRMEVLGCAQDIY 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	9474	76.3	2343	6	O18806 canis famil
2	9428	75.9	2343	6	O62730 canis famil
3	6276.5	50.6	2258	11	Q7TN96 rattus norv
4	3034.5	24.4	1639	13	Q804W6 fugu rubrip
5	2823.5	22.7	2119	13	Q90X47 brachydanio
6	2771.5	22.3	2183	11	Q88783 mus musculu
7	2685	21.6	1802	13	Q804W5 fugu rubrip
8	2511.5	20.2	1377	13	Q804X3 gallus gall
9	2343	18.9	1460	13	Q7SZN0 pseudonaja
10	1999.5	16.1	2102	11	Q7TPK2 rattus norv
11	1595.5	12.9	355	11	Q8BQ43 mus musculu
12	1433	11.7	745	13	Q804X4 gallus gall
13	1350	11.2	1156	11	Q80Y80 mus musculu
14	1389	11.2	1157	11	Q920Z4 mus musculu
15	1375	11.1	1157	11	Q920H8 rattus norv
16	1345	10.8	1156	4	Q9BQS7 homo sapien

17	1340	10.8	1158	4	Q9C058
18	1329	10.7	1087	13	Q7ZUI2
19	1308	10.5	1104	4	O75180
20	1299	10.5	1084	11	Q9TL97
21	1227.5	9.9	1048	6	Q9XT27
22	1124.5	9.1	847	11	Q8C4S2
23	1104	8.9	216	4	Q14286
24	932	7.5	626	13	Q90ZT2
25	911	7.3	407	13	Q9AYE0
26	740.5	6.0	1142	10	Q8LL91
27	700	5.6	503	11	Q8BV37
28	664	5.3	463	11	Q9RLX9
29	662.5	5.3	480	4	Q8N610
30	662.5	5.3	480	4	O43854
31	661	5.3	426	11	Q9WTS3
32	555.5	5.3	470	11	Q8C4U8
33	555.5	5.3	480	11	O8CBF7
34	555.5	5.3	480	11	O35474
35	555	4.5	312	4	O7Z3D2
36	514.5	4.1	363	6	O77718
37	488	3.9	335	4	Q9BTL9
38	469.5	3.8	901	4	Q9H2E4
39	469.5	3.8	901	4	Q9H2D5
40	469.5	3.8	906	4	Q9H2D4
41	469.5	3.8	906	4	Q9H2E3
42	469.5	3.8	931	4	Q7Z3T9
43	466.5	3.8	384	11	Q8C8K0
44	462.5	3.7	926	11	Q8QZY7
45	450.5	3.6	921	11	Q9QX38

ALIGNMENTS

RESULT 1

O18806	O18806	PRELIMINARY;	PRT; 2343 AA.
AC	O18806;		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Factor VIII.		
GN	F8.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_taxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RA	Cameron C., Notley C., Hoyle S., McGlynn L., Hough C., Kamisue S.,		
RA	Giles A., Lillicrap D.;		
RT	"The canine factor VIII cDNA and 5' flanking sequence."		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
CC	!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
DR	EMBL; AF016234; AAS87412.1; -.		
DR	HSSP; P00451; 1CFG.		
DR	GO; GO:0005507; F:copper ion binding; IEA.		
DR	GO; GO:0007155; P:cell adhesion; IEA.		
DR	InterPro; IPR001117; Cu-oxidase.		
DR	InterPro; IPR008972; Cupredoxin.		
DR	InterPro; IPR000421; FAS58 C.		
DR	InterPro; IPR008979; Gal_bind_like.		
DR	Pfam; PF00394; Cu-oxidase; 3.		
DR	Pfam; PF00754; F5_F8 type C; 2.		
DR	SMART; SM00231; FAS58_2.		
DR	PROSITE; PS01285; FAS58C.1; 2.		
DR	PROSITE; PS01286; FAS58C.2; 2.		
DR	PROSITE; PS00022; FAS58C.3; 2.		
DR	PROSITE; PS00079; MULTICOPPER OXIDASE1; 3.		
DR	SEQUENCE 2343 AA; 265829 MW; A854FAE571C3B399 CRC64;		
SQL			

Query Match

76.3%; Score 9474; DB 6; Length 2343;

Best Local Similarity		77.0%;	Pred. No. 0;
Matches 1814;		Conservative 197;	Mismatches 321; Indels 24; Gaps 12;
QY	1	MQTELSTCFELLCPLLCFCSATRRYYLGAVELSWDMQSD-IGELPVDVARPPRPVKPFPPF	59
DB	1	MOVELYTCCFLCLLPFLSATRKYYLGAVELSWDMYQSDLLSALHADTFSFSRVPGLPL	60
QY	60	NTSVVTKTLFVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVTITKNWASHPVSLHA	119
DB	61	TTSTVTRKTVFVEFTDHLFNIAKPPPMWGLLGPITQAEVVDVTVTIVJKNWASHPVSLHA	120
QY	120	VGSYVWKASGEGAEYDDQTSOREKEDDKVPFGGSHYVYVQVLKENGPMASDPLCTYSYLS	179
DB	121	VGSYVWKASGEGAEYDDQTSOREKEDDKVPFGGSHYVYVQVLKENGPMASDPLCTYSYFS	180
QY	180	HVDLVKDLNSGLIGALLVCREGLAKETQTLHKFILLFAVDEDEKSHSETKNSLMQDR	239
DB	181	HVDLVKDLNSGLIGALLVCKESLAKERTQTQEFVLJFAVDEDEKSHSETNASLTQ--	238
QY	240	DAASARAWPMHTVNGVNGVESLPGLIGCHRKSVYWHVIGMGFTTPVHISFIEGHTFLVRN	299
DB	239	---AQAQCHLHATNGVARSUPLTVCHRSVYVWHVIGMGTTPVHISFIEGHTFLVRN	294
QY	300	HRQASLEISPIITLTATQTLMDLQGLLCSHSHQHDGMEAYVKVDSCEPEPOLIMQN	359
DB	295	HRQASLEISPIITLTATQTLMDLQGLLCHIPSHQHDGMEAYVKVDSCEPEPOLIMQN	354
QY	360	EEAEYDDYDTDSMDVVRDDNSFSFTQIRSVAKKHKTWWHYIAAEEBMDVAPLVL	419
DB	355	ED-KDYDDGYLSDMDVVSFDDSSSFTQIRSVAKKHKTWWHYIAAEEBMDVAPSGP	413
QY	420	APDDRSYKSYLNGPGORICRKYKVRFMAYTDETEKTREA-QHRSGLGLPLLYGEVGT	479
DB	414	TPDKRSHNYLNGPGORISKKYKVRFVAYTDETEKTREAQYRSGILGLPLLYGEVGT	473
QY	480	LLIIFKQASRPYNIYPHGITDVRPLYSRRLPGVKVHLKDFPLPGBITPKYKWTVTVDG	539
DB	474	LLIIFKQASRPYNIYPHGINYVTLHTGRLPGVKVHLKDMPLPGBITPKYKWTVTVDG	533
QY	540	PTKSOPCLTRYSSFVNMERDLASGLIGPLLICYKESVDQKGNQIMSKRNVILFSVPD	599
DB	534	PTKSDPCLTRYSSFINLFRDLASGLIGPLLICYKESVDQKGNQIMSKRNVILFSVPD	593
QY	600	ENRSWYLTENIQRTLPNPAGVQLDEPFOASNTWHSINGVDFDSLQSLCHEVAYWYL	659
DB	594	ENRSWYLTENNORLEPMDVAVQPHDPFQSLNIMHSINGVDFDNQSLCHEVAYWYL	653
QY	660	SIGAQTDFLSYFPGGYTPKRWYVEDTLTPFSGETVFMMSMNGLIWLGCHNSDFNR	719
DB	654	SVGAQTDFLSYFPGGYTPKRWYVEDTLTPFSGETVFMMSMNGLIWLGCHNSDFNR	713
QY	720	GMTALLKYSSCKMTGYVEDSYEDI SAYILSKNAIETPRSFQNSRHPSTROKQFNT	779
DB	714	GMTALLKYSSCNRIIDDYEDYEDIPTPLNENNVNKRFSQNSRHPSTKQQLKAT	773
QY	780	TPENDIEKIDQSGERTQLIKAQSVSSDDLMLLQGPITPRGLFLSDLREATDR--ADDD	831
DB	774	TPENDIEKIDQSGERTQLIKAQSVSSDDLMLLQGPITPRGLFLSDLREATDR--ADDD	831
QY	840	SPGA-DSNNNS-SEMTHFPQLHSGDMVTPBESGQLALAEKIGTTAATTELKLDKPVSS	899
DB	832	SRGATERNKGPPEVASLAPELRHEDREFTPEPELQURLAENLGNTVTELKLDKLISS	851
QY	900	TNNLIS--TIESDNLAACTNTSLSGPPSMVHYDSOLDTTLTFCKXSSPLTSGGSL	957
DB	892	SBSLMTSTPTISDKLAAATEKTLGSLGPNMVSVPFNSHLGTVIVGNSSSHLIQSGVPLE	951
QY	958	SBENNDKLESLCMNSQESSMGKVSSTESGRIFKGAHGPALLTKDNALFKVSLSL	1017
DB	952	SEEDNDSKLELAPLXNIQSSSLRENVLSMESNRIFKEERIEGPNASLIKDNALFKVNISSV	1011
QY	1018	KTKNTSNNAATNRKTHIDGPSLLIENSSVQWNI-LESQDTFKKVTPLIHDRMLMDQNAT	1076

Dd	1012	KTNRAPVNLTNRKTRVTAIPTLLIENSTSVQWQIMLERNTEFEKVTSLHNSTFMDRNTT	1071
Qy	1077	ALRLNHSVSKTTSKNNMVMVQKKEGPIPPDAQNDMGFFXKALFLPSPARWIQTHGKNS	1136
Dd	1072	ALGLNHSVSKTTLSENVAHQKEDPVPLAENPDLSSXIPFLPD--WI-KTHGKNS	1127
Qy	1137	LSNGQGPSKQIVSGLPEKSVBGNFLSEKOKVUVVGKEFTKOVGLKEMVFPSSENFLIT	1196
Dd	1128	LSSEQRSPSKQTLISGSEKSVKDONFLGEE-KVVVGEDFTKDTSELQ-IFPNKNSIPFA	1185
Qy	1197	NLDNLHNTHNQEKKI0REIEKKETLLOENVVLPOIHTVGTCKNFMKNLFLLSRQWVE	1256
Dd	1186	NLANVOENDTYNQRKSPBEIERKEKLTQENVALPQAHMTGKNKFLNLFLLSTKQWVA	1245
Qy	1257	GSYDGAYPVLQDFRSLNDSTNRKTKHTAHFSKKEEENLEGLGNQTKQIVKVACTTRI	1316
Dd	1246	GLEBQPYTPILOTRSINDSPHSEGHIMANFSKIREEANLEGLGNQTNOMVERFPDSTRM	1305
Qy	1317	SPNYSQONFVQBSQKALIKQFRLPIEETEELEKRIIVDDTSTOWSKMKHLLFSTLTQIDY	1376
Dd	1306	SSNAS-CHVITQRGKSLQPRLSQGEIERFERKVIANDTSTOWSKMNYLAQGLTQILEY	1364
Qy	1377	NEKEKGAITOSPLSDCLTSHSITPOANRSPPLPIAKVSSPSPURPVLTRVLFDQDNSSHLP	1436
Dd	1365	NEKEKRAITOSPLSDCSMRNHVITQWDSALPVAKESASPSVRHTDLTKIPSOHNSHLP	1424
Qy	1437	AA----SYRKDKSGVOESSHFLOCAKKNLSLAILTLEMTGDQREVGSLGTSATNSVYTK	1492
Dd	1425	ASACNYTFRBETSGVQEGSHFLQEAKENNLSLAFVTLGITTEQQGFSSLGKSATNQPVYK	1484
Qy	1493	KVENTVLKPDLPKTSQKVELLPKXHHYOKDLPPTSTNSGSPGHLLDVEGSLLOQTEGAI	1552
Dd	1485	KLENTVLLQGLSETSKVLLSQWHDQSDSPXTSNDSPSGHLDLMGKIPLQMTQGPV	1544
Qy	1553	KWZANPBGKVPFLPRAVATESSATSBSKLLDPLAWDNHYGTQIPKEBWKQSKSPKTAFK	1612
Dd	1545	KXNTNSPGKVPFLKWAATESSEKIPSKLLGVLDWDNHYDTQIPSEWKQSKSQNTAFK	1604
Qy	1613	KKDTILSINACESNHAAINEGONKPEIEVTTWAKQGRTERLCSQNPVPLKXQREITRT	1672
Dd	1605	RKDITLPLGFCENNDSAAINEGDKPQREAMWAKQGPGRUCSQNPVPSKHQREITVT	1664
Qy	1673	TLOSDQBEIDYDITISVMKKEDPDIYDEDENSQSPRSFOKKTHTYIIAAVERLWDYGSS	1732
Dd	1665	TLQPEEDKFEYDDTFSIEMKREDFDIYGYDENQGLRSFOKKTHTYITAAVERLWDYGSR	1724
Qy	1733	SPHVLNRQAQSGSVPOPKKVFQFETDGSFTOPLYRGELNEHGLGLPGYIRAEVDENIMV	1792
Dd	1725	SPHILNRQAQSGDVQOPKVVVFOFSTDGSFTOPLYRGELNEHGLGLPGYIRAEVDENIVV	1784
Qy	1793	IFRNOASRPYFYSLSLISYEBDQOQAEPKKNFVKPNETKTYIFWKVOHHMAPTKDFFDCK	1852
Dd	1785	IFKQOASRPYFYSLSLISYDEDECGGAEPREKFNVPNETKIIFWKVOHHMAPTKDFFDCK	1844
Qy	1853	AWAYPSDVLDEKDVHSGLIGPLLVCHNTNLNPAHGRQVTVQEFALPFTIIEDETKSWYFTE	1912
Dd	1845	AWAYPSDVLDEKDVHSGLIGPLLVCRSNTLNPAHGRQVTVQEFALPFTIIEDETKSWYFTE	1904
Qy	1913	NMERNCRAPMIONEDPTFKENYFPAHATNGVIMDTLPLCMAOODORIRWYLLSMGNSNI	1972
Dd	1905	NZERNCRAPCNQVEDPTLKNFPAHATNGVYKDTLPLCVAQOQKVRWYLLSMGNSNI	1964
Qy	1973	HSIHFSGHVFTVRKKEEYKMAVNLNYPGVFTVEMLPSKAGIWEVECLIGEHLHAGMSTL	2032
Dd	1965	HSIHFSGHVFTVRKKEEYKMAVYNLYPGVFTVEMLPSQVGIWEI ECLIGEHLQAGMSTL	2024
Qy	2033	FLVYSNKOQTPLGMASGHIRDPOQTASQYQWAPKPLARLHYSGSINAWSTKPPFSWIKV	2092
Dd	2025	FLVYSKKOQTPLGMASGHIRDFOQTASQYQWAPKPLARLHYSGSINAWSTKPPFSWIKV	2084
Qy	2093	DLIAPMIHGIKTQARQKPFSSLYISQFLIMYSLDGKKWQYCRGNSGTCTLWFFGNVDSS	2152
Dd	2085	DILIAPMIHGIKTQARQKPFSSLYYSQFLIMYSLDGNKWSIRGNSGTCTLWFFGNVDSS	2144


```
QY 2153 GIKHNIENPPIIARIYIRLHPHYSIRSLRMLMGLCDLNSCSMPLGMSKASISDAQITAS 2212
DB 2145 GIKHNIENPPIIAQYIRLHPHYSIRSLRMLMGLCDFNCSMPLGMSKASISDAQITAS 2264
QY 2213 SYFTNVEATWSPSKARLHLOGSNAPVPOVNNPXELQVDFOKTMKVTVGTTOGVKSLLT 2272
DB 2205 SYLSSMIATWSPQARLHLOGRTNAPVPOVNNPXELQVDFOKTMKVTVGTTOGVKSLLI 2264
QY 2273 SMYKFEFLISSQGHOWTLFFQNGKVKVFOGNOISDFTFVVVNSIDPPLRLYLRHPOSW 2332
DB 2265 SMYKFEFLISSQGHENWTLFLQNGKVKVFOGNRDSTFVNRLEPPLVARYVRLHPOSW 2324
QY 2333 VHQAALMEVLGCSAQ 2348
DB 2325 AHHAALREVLGCDTQ 2340

RESULT 2
O62730 PRELIMINARY; PRT; 2343 AA.
ID O62730;
AC O62730;
DT 01-AUG-1998 (TREMELrel. 07, Created)
DT 01-AUG-1998 (TREMELrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE FactCr: VIII.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
XN (1)
XP SEQUENCE FROM N.A.
RC TISSUE=kidney, and spleen;
RA Gordy P.W., Bowen R.A.;
RT "Characterization of the canine factor VIII cDNA.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ Databases.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR BMBI; AF043489; RAC05384.1; -.
DR HSSP; P00451; 1CFG.
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0017155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; GalBind_like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8 type C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2.
DR PROSITE; PS00222; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
SQ SEQUENCE 2343 AA; 265613 MW; F612D744ADAADD99 CSC64;

Query Match 75.9%; Score 9428; DB 6; Length 2343;
Best Local Similarity 76.7%; Pred. No. 0;
Matches 1806; Conservative 200; Mismatches 326; Indels 24; Gaps 12;

QY 1 MQELSTOFPCLLRFCFSATRYIYGAVELSDYMQSD-IGELPVDARPPRPVPSFFP 59
DB 1 MQVELYTCFCLLPFSLSATRYIYGAVELSDYMQSDLLSALHADTSFSSRPVGSLLP 60
QY 60 NTSVYVKKTFVFTDHLFIAPRPVWMLGPTTQASVYDVTWITLKNMASHPSLHA 119
DB 61 TTSVYVKKTFVFTDHLFIAPRPVWMLGPTTQAEYDVIIVLKNMASHPSLHA 120
QY 120 VGSYVXASGAEDDQTSOREKEDDKVFPFGGSHYVQVLKENGPNWASDPLCLITYSL 179
DB 121 VGSYVXASGAEDDQTSOREKEDDNVIFGESHYVQVLKENGPNWASDPLCLITYSFP 180
QY 180 HVDLVKDLNSGLIGALVCFRGLAKSKTQTLHKFILLFVAFVDEGKSWHSETKNSLMQDR 239
DB 181 HVDLVKDLNSGLIGALLVCKEGLSAKERTQTLQFVILLFVAFVDEGKSWHSETNASLTQ-- 238
```

```
QY 240 DAASARAWPMHTVNGYVNRSLPGLICCHRSKYVHWIGMGTTPPEVHSIFLEGHTFLVRN 299
DB 239 ----AEAQHELHTINGVNRSLPGLTVCHRSKYVHWIGMGTTPPEVHSIFLEGHTFLVGN 294
QY 300 HRQASLEISITPFTAQTLMDLGOFLLSCHLSSHQHDGMEAYKVVDSCPEEPQLMKNN 359
DB 295 HRQASLEISITPFTAQTLMDLGOFLLECHPSHQHDGMEAYKVVDSCPEEPQLMKNN 354
QY 360 EEAEDYDDDLTDSMDVVRFDNNSPSFIOIRSVAKKHPTKTVHWYIAAEBEDWDYAPLVL 419
DB 355 ED-KOYDDGLYGSMDVVSFDDSSSPFIOIRSVAKKHPTKTVHWYIAAEBEDWDYASGP 413
QY 420 APDDRSYKSOVLNNGPQIRIGRYKKVRFMAYTDETFKTREAIQESGILGFLLYGYGDT 479
DB 414 TPNDRSKHNLYLNGPQIRIGRYKKVRFMAYTDETFKTREAIQESGILGFLLYGYGDT 473
QY 480 LLIIFKNOASRPNIYPHGIDTDEPLYSRRLPKVKHLKOPPLPGLPGIEKYKMTVTVEDG 539
DB 474 LLIIFKNOASRPNIYPHGINYVTPHTGRLPKVKHLKOPPLPGLPGIEKYKMTVTVEDG 533
QY 540 PTKSDPRCLTRYSSFVNMBERLASGLIGPLLI CYKESVDQKGNQIMSKDNVILFSYVD 599
DB 534 PTKSDPRCLTRYSSFINLERDALASGLIGPLLI CYKESVDQKGNQIMSKDNVILFSYVD 593
QY 600 ENRSWYLTENIOTFLPNPAGVQLEDPPFQASINMHSINGYVDFSLQLSVCHEVAYWYL 659
DB 594 ENRSWYLTEDMQRTLPNADVVQPHDPFQLSNTMHSINGYVDFNLQLSVCHEVAYWYL 653
QY 660 SIGAQTDFLSVFFSGYIFKHWVYEDTLTPFSGSETVFMSENPGLMILGCHNSDFNR 719
DB 654 SVGAQTDFLSVFFSGYIFKHWVYEDTLTPFSGSETVFMSENPGLMILGCHNSDFNR 713
QY 720 GMTALLKVSQCDKGTGYEDSYEDISAYLLSKNNAIEFRFSQNSRHPSTROKQFNATT 779
DB 714 GMTALLKVSQCNRIIDDYETIEDIPPLANENNVIKPSFSQNSRHPSTKEKQKATT 773
QY 780 IPENDIEKTDPMFAHRTMPKIQNVSSDILLMLRQSTPHGLSLSDLOEAKYETFSDDP 839
DB 774 IPENDIEKIDLQSGERTQILKAQSVSSDILLMLLQGNPTPRGLFSLDREATDR--ADQH 831
QY 840 SPGALDSNNSENTHFRPOLHSGDMVTPESGLOLRNEKLGTTAATLKKLDFKVSS 899
DB 832 SFGAIERKNGPPEVASLRPELHSHSDBRFTPPPELQRLNENLGNVTVELKKLKLKISS 891
QY 900 TSNLLIS--TIPSDNLAAGTDNTSIGPPSMVHYDQSDTLTGCKKSPLETSGGSL 957
DB 892 SSDSLMTSPTIPSDKLAATEKTGSLGPNMVSFHNGHLGTVFGNNSHLLIQSVPL 951
QY 958 SEENDSKLLESGLMNSQESSWGKQVSTESGRLPKGKAHGPALLTKDNALPKYSISLL 1017
DB 952 SEENDSKLLEAPLMNIQESSLRENVLMSNRLFKERIRGPASLIKONALPKVNISS 1011
QY 1018 KTKTKSNRSATNRKTHIDGPSILLIENSPSVQNI-LESDETFKVTPLIHDRMLMDKNAT 1076
DB 1012 KTNRAPVNTTKRUTKVAIPILLIENSTSVQDILMLRTEFEKVTSLIHNTFMDRNTT 1071
QY 1077 ALRLHMSNKTSSKNMVMVQKKEGPIPPDAQNDPMSFFKMLFPEBSARWTQTHGKNS 1136
DB 1072 ALGLNHSNKTLSKNVEMAHQKEDPVLRAENPDLSKKIPFLPD---WI-KTHGKNS 1127
QY 1137 LNSGQSPSKQLVSGPEKSVGQNFSLSEKNKVYVCKGFTKDVGLKEMVFPSSRNLELT 1196
DB 1128 LSSEQSPSKQLVSGPEKSVGQNFSLSEK-KVVGEDEFTKDETEQCE-IFPNNSKIFFA 1185
QY 1197 NLDNLEHNTNQEKIKQEEIEKKEITLIQENVVLPOIHTVTGKNPMKMLFLISTQNVE 1256
DB 1186 NLANVQENNTYQEKKSLEIERKEKLTQENVALPOAHTMIGTKNPKLNLFLISTQNVA 1245
QY 1257 GSYDGAAPVLQDFRSLNDSTNRKHTAFBFGKKGEENLEGNGQTKQIVKAYACTRI 1316
DB 1246 GLEEQYTFILQDTRSLNDSPHSEGHMFANFSKIREANLEGLNGQTNQWVERFPSTTR 1305
```

```
QY 1317 SPNTSQNEVTCRKBALQKQSLPLEETLEKRIIVDDTSTQWKNMKHLTPSTLTQIDY 1376
Db 1306 SSNAS-QHVIITQGRKSLUKQRLQGEIKFERKVIANDISTQWKNWNYLAQGLTQIEY 1364
QY 1377 NEKEKGATITQSPSLDCLTRESHISIPQANRSPFIKAVSPFPIRPIYLTRVLFDQNSHLP 1436
Db 1365 NEKEKGATITQSPSLDCLTRESHISIPQANRSPFIKAVSPFPIRPIYLTRVLFDQNSHLP 1424
QY 1437 AA----SVRKQDQGVQBSHFLQAKKNLSIALITLQNTGQREVSLGHSATNSVTYK 1492
Db 1425 ASACNYTPRRTSGYQEGSHFLQEAQRNLSLAFVTLGITGQCKFSGLKGSATNQPMYK 1484
QY 1493 KVENTVLPKPDLPKTSYGVEILLPKVHIYQKOLFPETSGSPGHLDLVEGSLAQTEGAI 1552
Db 1485 KLENTVLLQPLGSETSDVQVLLSQVEVQBSFPFKTSNDSPGHLDLNGKFLPKTQGFV 1544
QY 1553 KWNENRPFKVPFLAVATESSAKTSKLLDPLAWNHYGTQIPKEBWSQKSPKTAFK 1612
Db 1545 KWKNTSPGKVFLLKWATESSEKIPSKLLGLVLDNHYDTQIPSEWKSKQSQTNTAFK 1604
QY 1613 KXDTILSLNACHSNHAIKAINBQNKPLEVTVAKQGRTERLCSNPVLPKHOEIEITET 1672
Db 1605 EXDTTLPLGPCNNDSIAINEGDKPOKQKPEAKWAKQGEPRLCSONPPVSRHHQREIIVT 1664
QY 1673 TLQSDQEEIDYDTISVEMKKEDFDIYDEDNQSPESFOKTRHYFIAAVERLWDYGMSS 1732
Db 1665 TLQPEDKEFYDDTSIENKREDFDIYDVEQGLRSQKKTREHYFIAAVERLWDYGMSS 1724
QY 1733 SPHVLRNRAQSGVQPKVQVQETDGSFTQPLIRGELNEHGLGLGPVIRAEVDENIMV 1792
Db 1725 SPHVLRNRAQSGVQPKVQVQETDGSFTQPLIRGELNEHGLGLGPVIRAEVDENIMV 1784
QY 1793 TPKNQASRYSVYSSLIISVEEDQKQKPRKFNENETKYFWKVOHHMAVTKQEDFCK 1852
Db 1785 TPKNQASRYSVYSSLIISVEEDQKQKPRKFNENETKYFWKVOHHMAVTKQEDFCK 1844
QY 1853 AWAYFSDVLEKDVHSLGILPVLVCHTNTLNPAHGRQVTVQBFALFTTFIDETKSYFTE 1912
Db 1845 AWAYFSDVLEKDVHSLGILPVLVCHTNTLNPAHGRQVTVQBFALFTTFIDETKSYFTE 1904
QY 1913 NMERNCRACNTQMDPTTKENVRFAINGYIMDTLPGLVMAQDQIRYLLSMGSENI 1972
Db 1905 NLERNCRACNVQKEDPTLKNFRFAINGYIKDTLPGLVMAQDQKVRWYLLSMGSENI 1964
QY 1973 HSHPGSHVFTVRKKEEYKMAVNLVPGVFPETVEMLPKSKAGIWRVCLIGEHLAGMSTL 2032
Db 1965 HSHPGSHVFTVRKKEEYKMAVNLVPGVFPETVEMLPKSKAGIWRVCLIGEHLAGMSTL 2024
QY 2033 FLYVSNKQOTPLGMASSHIRDFQITASGOYQWAPKLAELHYSGSINAWSTKBPFWIKV 2092
Db 2025 FLYVSNKQOTPLGMASSHIRDFQITASGOYQWAPKLAELHYSGSINAWSTKBPFWIKV 2084
QY 2093 DLLAPMIHGIKTQCAROKESSLIYSQFIIMYSLDGKKWCTYRGSNTGLMWFFGNVDS 2152
Db 2085 DLLAPMIHGIKTQCAROKESSLIYSQFIIMYSLDGKKWCTYRGSNTGLMWFFGNVDS 2144
QY 2153 GIKHNIFNPPIIARYIRLHPTHYSIRSLRWELMGCDLNSCMPLGMESKAIISDAQITAS 2212
Db 2145 GIKHNIFNPPIIARYIRLHPTHYSIRSLRWELMGCDLNSCMPLGMESKAIISDAQITAS 2204
QY 2213 SYFTNMFATWSPKARLHLQGRNAMPQVNNKPEWLOVDFOKTMKVGTGTTQGVKSLIT 2272
Db 2205 SYLSSMLATWSPQARLHLQGRNAMPQVNNKPEWLOVDFOKTMKVGTGTTQGVKSLIT 2264
QY 2273 SMYVKEFLISSQDGHQWTLFPQNGKVKVFGQOGSETPVWNSLDPLLTRYLRHPOSW 2332
Db 2265 SMYVKEFLISSQDGHQWTLFPQNGKVKVFGQOGSETPVWNSLDPLLTRYLRHPOSW 2324
QY 2333 VHOIALRMEVLGCEAQ 2348
Db 2325 AHIIALRLEVLGCDTCQ 2340
```

```
RESULT 3
Q7TN96 PRELIMINARY; PRT; 2258 AA.
AC Q7TN96;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Factor VIII.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Liver;
RA Katzka M., Geisen C., Seifried E., Oldenburg J.;
RT "Sequence of the rat factor VIII cDNA."
RL Submitted (AUG-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY362193; AAQ21580.1; -.
SQ SEQUENCE 2258 AA; 251299 MW; 13AF91C789059B1D CRC64;

Query Match 50.6%; Score 6276.5; DB 11; Length 2258;
Best Local Similarity 53.2%; Pred. No. 0;
Matches 1271; Conservative 317; Mismatches 630; Indels 171; Gaps 29;

QY 1 MQIELSTCFFLCLLRFCFSATRRYVLGAVELSWDMQSDL-GELPVDARFPRPVKSPFF 59
Db 1 MEAARGLCFFLSLCAIASCATKRYLGAVELPMDYVSGASGACRSDSRFPFGTPAP--- 57
QY 60 NTSVYVKTLFWETDHLFNIAKRPWMLGGLPTIQAEVYDVTVTILKNASHVPSLHA 119
Db 58 SARVHTKTVFVYMDRPFHTAKPRPLWMLGGLPTITWTEVHDVTVTILKNASHVPSLHA 117
QY 120 VGVSVMKASGAEYDDQTSQREKEDKVPFGSHTYVMQVLKENGPMASDPLCLTYSYL 179
Db 118 VGMSFWKASGAEYDDHSSPAEKDDKVLPGESHTYVMQVLGSGPMASDDEPCLTAYLS 177
QY 180 HVDLVKDLNSGLIGALLVCRGSLAKRKTQTLHKIILLPAVFDEGSKSHSETKNSLMQDR 239
Db 178 HVDLVKDLNSGLIGALLVCRGSLAKRKTQTLHKIILLPAVFDEGSKSHSETKNSLMQDR 230
QY 240 DAASARAPKXHTVNGYVNNRSLPGILGCHKSVYHVTGCTTPEVHSITLHGHTFLVRN 299
Db 231 DEASTAQPTVHTVNGYVNNRSLPGILGCHKSVYHVTGCTTPEVHSITLHGHTFLVRN 290
QY 300 HQASLEISPTFLAQTLLMDLQGLLSCHTSSHQHDMGEAYKVVDSCPEPQIMKN 359
Db 291 HQASLEISPTFLAQTLLMDLQGLLSCHTSSHQHDMGEAYKVVDSCPEPQIMKN 350
QY 360 EBAEYDDDLTSEMMDVVRFDNNSPFTQIRSVAKKPKTWHTYIAAREEDWDYAPVL 419
Db 351 EBTEDYDD--LDSEMDVFTWDPDAAP-FVQVRSAAKRHPKTVHTYIAAREEDWDYAPAA 407
QY 420 APDGRSKQYLNNGPQIRGRYKVRPMAYTDETPKTRREALQHSIGLPLLYGEVGT 479
Db 408 TLDEGSLRNYLGRGPRIGSKYKVRVYAYIDGTFHTREATPREAGLGLPLLYGEVGS 467
QY 480 LLTIIFNQASRPNYIYPHGITDVRPLYSRRLPGVKHLDKDFPLPGEIFKYKWTVTVEG 539
Db 468 LLTIIFNQASRPNYIYPHGIRDVGAHVAGRLPRGVKHXDLPIRPGETPKYKWTLTAE 527
QY 540 PKSDPRLCTRYSSPVNMRDLASGLIGFLICVKESVDQKGNQTMDSKRNVLFSVFD 599
Db 528 PARSDARCVTRYASAVDPERDLASGLIGFLICVKESVDQKGNQTMDSKRNVLFSVFD 587
QY 600 ERESWYLTENIORPLPNPAGVQLEDPEFOASNMHSINGVYFDSLSQSLCHEVAYWYL 659
Db 588 ENKSWTISNMRARFLPDBAHVQLQDEFEASNMHSINGVYFDSLSQSLCHEVAYWYL 647
QY 660 SIGAQTDLSVFFSGYTFKEKMYEDTLTFPSGETVFMNMPGLWILGCHNSDFNR 719
Db 648 SVGAQTDLSVFFSGHTFRRAVYADTLTVFPHSGVTVFWSMNDNPGVWVILGCHNPFRES 707
```

QY 720 GWTALLKVSCKNTGDIYEDSYEDISAYLLSKNNALPRSFQNSRHPSTROKQNFATT 779
Db 708 GWTALLKVSCKDGVSDYIYGETYGVJA-PLANDNAVDPSPFQNSNHLHRKKS--- 763
QY 780 IPENDIEKTDWFAPHTPMPKIQNVSSDILMLRSGPTGHLGSLDLQAKAYETSDDP 839
Db 764 -----TEM-----RPLSHPTQHGSPWSDSQGDVDAVHRAE 795
QY 840 SPGADSNNSLSLSEWTHFRPOLHSGDMVFTBESGLQRLNEKLGTTAATLKLDFKVS 899
Db 796 SPAAHNEGPAEAAQLEPKHPGKXALPFS---WLRKRSLLVTTWEAKKKLJLQVQ- 851
QY 900 TSNLLSTIPSDNLAAGTDDNTSSILGPPSMPV-----HYDSQLDTTLFGKSSSLTE 950
Db 852 -----VSGLPD-----DRTAVVA PDDPMAACKAGSGFPDESSPAALGKKMYPRIR 899
QY 951 SGGPLSSENNDSKLLLEGLMNSQESWKNVSTESGRLPKGRKRAHGALLTKDNALF 1010
Db 900 PHGPLSLIEGNRDSNSDSTLMYRLGSPGDATSWTENRRGLGKSRHVAFLARGNTLL 959
QY 1011 KVSISLLKTKNTSNNATNKTTHIDGSLLIENSFVWQILLESDFKFKVTPLIHDSML 1070
Db 960 -----SDVDEKSHAPGPTS VGNSTAAVQDITLCTCSILEVTPPIHRIIL 1004
QY 1071 MDKNATLRNHNMSKNTTSKKNMEMVQKKEGPPIPPDAQNPDMMFFKMLFLPESARMIQR 1130
Db 1005 SDKSATYLRPHRTPDRTITSTERKDIRHEKGDVLPQDADTSAPFSEAPFLSESTDLKE 1064
QY 1131 THGNKNSNGSGGPPKQIVSLGPKSVSEGGNFKSEKKNVVVGKEPFTKDVGLKEMVFPSS 1190
Db 1065 ANGNSAKPEQEPSEFQGLVIMYIKCMENOSFSSEKKNIAQGGFTKNTGLEDTVPFRK 1124
QY 1191 RNLFLTLNLDLHNENTNOEKKIOEETBEKKTLLQENVLPHHTVGTCKNFMKNLFLLS 1250
Db 1125 TSVELTIVAKQESGRHNCB-NIPQAVTEKAPLEKFGALPQVHIATGSKNPPDMFVLG 1193
QY 1251 TRQNVESYDGAYAPVLDQFSLNDSNTNRKKTHTAFHSKKEES--NLEGLNGTKOIVE 1308
Db 1184 TGQNI-NLHEETVYVHKVNPWTNPDTTRQIPMVHFFKKEBETNSGGLVKNKTREIVR 1242
QY 1309 KYACTTISNPTSQNFVQSKKALQFRLPLETELEKRIIVDDTSTQWKNMKHLTP 1368
Db 1243 NY-----PSQKNSVARREQASGRK-----ASARLWPDVNRSIQ 1277
QY 1369 STLQIDYNEKKGATQSPSLDCLTRHSPQANRSLPIAKVSSPSIESPIYLVTLVLF 1428
Db 1278 SLLKQIDHRKGRKFFIESRRADS-SGTSTQTNHSPSHVVKMSAFP---PRDIRIBS 1333
QY 1429 QDNSSHLPAASY-----RKDSGVQESSHFLOCAKKNLSLAILTLEMTGDQREVSLGTS 1484
Db 1334 RD-SSQVNVSSYAYDFETGSSRIRESSPFLNETETDNPDLAVPPRFRVRRGRZASPEKV 1392
QY 1485 ATNSVTYKKVNTVLPKDLPLKTSKVELLPKVHIYQKDLPTETNSGSPGHLLVBSGL 1544
Db 1393 NTHSAPCDKPNVSVSEFVSPCEAVTALPHVSTQBEELPAGSSCBRRGHVDLFLBIS 1452
QY 1545 LQCTEGATKMEANRPGKVPFLRVATESSAKTPSKLLDPLAWDNHYCQIPKEBWKQBX 1604
Db 1453 LQRTGPGVERGRGRSG-----DTEGHTENPGKTPSPL-----PQMPKQDQSGSAG 1499
QY 1605 SP-----EKTAFKKKOTILSINACESHAAINEGQNKPEIEVTWAKQGRTERLCSON 1658
Db 1500 YPKICRWEDGTVLPBRPHNLSLGA-----KKKNLPRREATRVQGEAPSLVSPE 1549
QY 1659 PVLKRGHOREITR-----TTLOSQDEEI--DYDDTILSVMKKEDFDIYDE 1701
Db 1550 PLVLRERRPREASTLLPGGEMRBDGCVTAEYDDDDDVMAEYDDDAVTVDTPEDCDIYGE 1609
QY 1702 DENQSFERSFOKTRHYFTAABERLDYGVSSSPHVLNRASQGSVPQPKVVVQBFDTGS 1761
Db 1610 EAGQCPRGFOQKTRHYFIAABELMDYGVVSPGALGDAWSGDARFKVVRFTTGS 1669
QY 1762 FTQPLYRGELNHLGLLOFYIRAEVEDNMVTRNQASRPYSFYSLISYEEDQSQGABP 1821

Db 1670 FTQVRHGSLDARLGLLGPYIRAEVEDNMVTRNQASRPYSFYSLISVSPYEAAD-GGAAP 1728
QY 1822 RKNFVKVNETKITVFWKQVHMAPTKDFDCKAWAYFSDVDLEKDXHSGSLGILLVCHTNT 1881
Db 1729 RSNFVFPNETKITVFWVRPHMAPTKDFDCKAWAYFSDVDLERDLHSGSLGILLVCHAST 1768
QY 1882 LNPAGHRQVTVQBFALFFTIIPDETCKSWYFETENMERNCRAPCNITQMBEDPTPKENYRFAIN 1941
Db 1789 LHSSEARPLAVQBFALLFAVDFETCKSWYCAENLDRKCRPPRGTPAGDPAMREBYRFAVN 1848
QY 1942 GYTMOTLPLGLVMAQDQIRWYLLSYGNSNENIHSPSGHVFTVRKKHXYKALYNLYPGV 2001
Db 1849 GYVADALLGLVMAEGRRTRWHLHLSMGDPGHAQSVHFSVTVRVDGGEHRTAVCNLYPGV 1908
QY 2002 FETVEMLPKAGWVRVVECLLGEHLHAGMSTLFLVYNSKNKCTPLGMASGHIRDFQITASGQ 2061
Db 1909 FTTVEMLPSPAGIWRVSVLGEHLRAGMSALFLVYSTROQVPLMGASGYIRDSQITASGY 1968
QY 2062 YGOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTQGARQKPFSSLYISQFI 2121
Db 1969 HGLWTPQLARLHNAGSVNWSAKPEPAWIKVDLLAPMLHGIETQGARHRLSSLYVSQFI 2028
QY 2122 IMYSLGCKWQTVYKNSGTFLAVFFGNVDSGSKHNIIPNPTIARYIRLPHPTHYSRSTL 2181
Db 2029 IMYSLGCKWQVLYKNSGTSLAVFFGNVDASTVRHNRFPDPIVARVIRVHPTASRTAL 2088
QY 2182 RMELMGCDLNSCMPZGMEKSAISDAQITASSYFTNMFATMSPSKARLHQGRSNARWQP 2241
Db 2089 RMELMGCDLNSCASPAGNESGWSIDQVTASSHLSITFSKAPFALARLHLRGGNARWQP 2148
QY 2242 VKNPKWQLQVDFOKTKMVTGVTTQGVKSLTSMYKVEFLISSQDGHQWTLPTQNGKVY 2301
Db 2149 VNDPTQWQLQVLDQRTVAVTGVVQTQARSLLTAMFVKFLVSTQDGRHWTHTVLQDGKVY 2208
QY 2302 FQGNQDSFTTVVNSLDPPLLTRYLIRHPQSVWVHQIALRMEVLGCEAQL 2350
Db 2209 FQGNRDASTPMVNSLHPFRFTRYLRHPQVWERQIALRLEILGCEAQL 2257
RESULT 4
Q804W6 PRELIMINARY; PRT: 1639 AA.
AC Q804W6; (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Coagulation factor VIII precursor.
GN F8.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Ial K., Snell P., Elgar G.,
RA Tuddenham E.G.D., McVey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
coagulation genes from Gallus gallus and Fugu rubripes";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF465279; AA033374.1; -
DR GO; GO:0005507; F: copper ion binding; IEA.
DR GO; GO:0007155; P: cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 2.
DR Pfam; PF00754; F5 F8 type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.


```

1296 LVCTVGEISLQWRAKLLVNFQCSPLGKSGRICDSOIKASDYIGNLPLHARLDQS 1355
QY 2076 GSNIAWSTKEPFWIKVLDLAPMIHGIKTQGAQKQFESSYISQFIIMYSLDGKQWQYR 2135
Db 1356 GYINAWGYNEKSWIQVDLQAPTLHRVQIQGVRSNLRNNYITAFYVSIDQETWSTYR 1415
QY 2136 G-----NSTGTLNVFQGVNDSSGIKENIENPPIIARYVIRLPHYTHYIRSTLWELMGCD 2189
Db 1416 GSGSSRSSSSSTAKVFNGLDNRVKNPFPVPPVARYIRIHPYIYNORALRNELLGCD 1475
QY 2190 LNSCSMPLGMEKSAISDAQITASSYFTNMFATWSPSKARLHLQGRSNARPPQVNNPKEWL 2249
Db 1476 LNSCSPLGLQDRIPRESFVASSYSWSLRSWTFPSLARLHQEGSANARPPKNNPHEWL 1535
QY 2250 QVDFQKTMKVGTQGVQKSLTSMYKVEFLISSQDQHQWTLFFQNG--KVKVFQGNQD 2307
Db 1536 QVDLGVKRVITGVVTOGARSILTRNMVTFERSVTISRQDQAMSSVLEGSSOREKIFQGNND 1595
QY 2308 SFTPVNSLDPLTRYLRIHPQSVWVHQIALAMEVLGCEAQ 2348
Db 1596 SDEALTFDAPLFGVYIRIHPGLWINDIALRLVLGCDTQ 1636

RESULT 5
Q90X47 PRELIMINARY; PRT: 2119 AA.
AC Q90X47;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE SC:b22015.3 (Novel protein similar to vertebrate coagulation factor V
DE and VIII).
GN SC:B22015.3.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN SEQUENCE FROM N.A.
RA Lloyd D.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; AL590146; CAC94896.1; -.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0007155; F:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal_Bind_like.
DR Pfam; PF00394; Cu-oxidase; 2..
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS00022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE; 3.
SQ SEQUENCE 2119 AA; 240643 MW; DC0E806FFA8761E6 CRC64;

Query Match
Best Local Similarity 30.3%; Pred. No. 2.2e-159;
Matches 752; Conservative 379; Mismatches 827; Indels 525; Gaps 68;

QY 11 LCLRF-CFSAT---RRYVLGAVELSDWYNQSDGLGELFVDARPPRPVPSFFNFSVYVK 66
Db 13 LALLAEFCATAVERHYVIAAVNINWDTSGQQ-----RTQGSYK 53
QY 67 KILVFPTDHLFNIAKPRPFWMLGPTCAEYVDVITWILKNWASHPYSLEHVGYSYWK 126
Db 54 KVIYREYNEG-FKQPKAHLPSLGLLFTURGGQDIIIVTFRNADHPCLSPHFGIAYGK 112
QY 127 ASGCAEYDQQTQSREXEDDKVFPGGSHYVWQVLKENGPMASDPLCLITYSLSHVDLVXD 186

```

```

113 QSSGSLVFNTSLLEKNDVIOQGESEHTYQWDVTSVDTPTAAPPCCITTYLGHFDIVRD 172
QY 187 LNSGLTGALLVCHREGSLAKKKTQTLH---KFILLFAVEDSGKSWHSETKNSLMDQRDAAS 243
Db 173 YNTGLIGEMLIICKGTLDSSGNO-IHFHQESVLLFGVFDENKSWYS-----TGDS 221
QY 244 ARAPKMHVTVNGVNRSLPGLIGCHRSVYVHVIGMTTPEVHSIFLEGHTFLVRNHRQA 303
Db 222 PQPLNVKTYTINGYNGSVPLDICAHSKVSWHLGMSSEPELFSVHFNGQVLLHDGKTS 281
QY 304 SLBISPTFTTAOTLLMDIQOFLLSCHISHQHDGMEAYKVDSCPEPOLMKNNEAE 363
Db 282 AVGLISGTSATSASMTGVHPGRWLVSSSHISKHLAAGLHGYLNIRKC----- 326
QY 364 DYDDDLTDSMDVVRDDNSPSFIOIQRSAKVKHPTWVHYIAAEERDWDYAPILVLPDD 423
Db 327 -----DEYAPK-RLRTIQKESQWYTYMAAEVINDYAPNKPENMD 369
QY 424 RSVKSOYLNNQPORIGRYKKVRPMAYTDTFTKTR---EAIQHESGILGPLLYGEVDTL 480
Db 370 GDFESKYLKQGPQRIKGYKKAFTQYKDGWFKERABDKQKRGELGILGPVIRAQIRDI 429
QY 481 LIIPKQASRPYNYPHGITDVRPLYSRRLPKGVKHLKDPILPCEIFKYKWTVTVDGP 540
Db 430 KIVFNKASRPYSIYPHGLTIDKAAEGASYPQGN--QYTSVQGEITYTYTWSVEEDVP 487
QY 541 TKSDPRGLTRYSSFVNMRDLASGLIPLLICYKESVDQKNGQIMSDKRNVLIFSVEDE 600
Db 488 TDSDFRLTRMYHSADVADPRDIASGLVGPLLICQSLNKKNVQLKADKEQHAFMTVDE 547
QY 601 NRSYLATENIORFLNPAGVQLEDPFQASNMHSINGYVDSLQ--LSVCLHEVAYWYL 659
Db 548 NKSVMYQDENINTYCSDPKVKKDDPFEYKSNVMTINGYVYESQGLGFCHEIVTWEVS 607
QY 660 SIGAQTFPLSVFPGYTFKHMVYEDTLTLPFSGSETVFMSEWPGWILCHNSDPRNR 719
Db 608 SVGEQDVIQTATFYGHTPELKNREEDILSLPMTGETITMNVNIGIWMLASLNSHDSK 667
QY 720 GMTALLKVSCKNTGDY--YEDSVEDI-SAYLLSKNAIEPRFSQNSRHPSTROKFNAT 778
Db 668 GMVKFKDLECFR---DYVIDYED-----GKFTAWKPPTI----- 701
QY 779 TIPENDIEKTPWFARHTPMPKIQNVSSDILLMLLQSPTPHGLSLDLQAKYETPDD 838
Db 702 ---NEIKKEEPRAR----- 713
QY 839 PSPGAI DNSNSISEMTHFRPOLHSGDMVTFPESGLQLRLNEKLTGTTAATLKLDFKVS 898
Db 714 --PQVVEYSDLFZET-----LNRFTNNVKDEVEIIDTLFL 748
QY 899 STSNLI STIPSDNLAAGTDNTSSILGPPSPMPVHYDSQLDTTIFGKKSPLTESGGPLSL 958
Db 743 DODDGLLPIVEKSLGSSNEN-----LHNATLQSFIEHGO----- 783
QY 959 BENNDSKILLEGSLMNSQSSWCKNVSTESGRLFKGKAHGPALLTKNALFKVSIILK 1018
Db 784 -----LMEEGLDKGESSKNVLNDSTK-----ALLE 811
QY 1019 TNKTSNNS---ATNRKTH---IDGPSL--LIENSPSVQNILESDFEKKVPTLHIDRML 1070
Db 812 TTTTFDSNRVVALNNETSIILDFPIVERKVRSAFSKPMNEPEFSTMMFKITEHI----- 866
QY 1071 MDKNATARLNMNSKNTTSSKNMVMVQKKEGPIPP-DAQNPDMSF-----FKMLFLPES 1124
Db 867 ---NSSLERINAI:SPITETNINTWETHTDPSITPFDGSGTGMFTLEDDTALLNSSES 923
QY 1125 ARWIOETHGKNSLNSGQSPSPQLVSLGPKSVBQGNF-----LSEKNVVGKGE 1175
Db 924 EPLQNSONSENRIAFQEBELNAKDGTDVDSNNVSKNQIFKYNVPSGDTLSSNSKIQV---- 979
QY 1176 FTKDVLGKEMVFPSSRNFLTNLDNLHNNTNQEKIOEBIEKKEKTEKLEQNVLPQIHT 1235

```

```
Db 980 -----BEDFVLLDSSYFSBMST-----TMEYDVQKDTVKGESKETAQSQE 1020
Qy 1236 VTGKNFMKNLPLIISTRONGVEGSDGAYAPVLQD-----FRSLN-----DGTN 1278
Db 1021 LSSTKTKYSGEILLESPLDITSAFNLNDSVLRNNLSNRSNETLFWSSNAFSDSTN 1080
Qy 1279 RTKKT-----AHFKKG-EEENLEGLNQTKQIYEKVIACITRISNTSQNFVQSRK 1332
Db 1081 ATSSDSSATFADFNTFSNATPSDFNRISQMSDSSNATLSDSSNAT-----LSDSNA 1136
Qy 1333 ALKQFRLPLETELKRIIVDTSTQWKNMHLTPS-----TLQIDYN-- 1377
Db 1137 TLSD-----SSNATLSDSSNATLSDSSNATPSDSSNATPSDSSNATLSDSSNAT 1186
Qy 1378 -----EKKGAITQSPSLDCLTRSHSIPOANRSPPIAXVSPPSIRPIVLRVLFQ 1429
Db 1187 TTLSFSLSEMTYILSSANDTI-KSHSEVVSNTSQL-----SSSESTENISL---LYG 1237
Qy 1430 DNSSHLPAASYRKSGVQESSEPTLQAKXNNLSLAILFLEMTGQDQREVSLGTSATNSV 1489
Db 1238 SLN-----ASSMKNDSESESEBEVVIYNKH-SEAILTSHL--DOKE-EHWGYESKHEL 1288
Qy 1490 TYKKVENTV-----LKPDPKPTSGKV--ELLPKVHI-----YQKDLFTEISN 1531
Db 1289 VHEKLPDHVKNKYVKDKSAANSKPKLEKVKYQKVKPKGYKGMKTKSKDYKQPRSS 1348
Qy 1532 GSPGHLDLVEGSLLOQTEGAIKWNANRPGKVPFLRVATSS-AKTPSKMLDLPLAWDHY 1590
Db 1349 FSP-----RGFGSVLTGRSRP-----VSSSEDELTEKPIVIGVPRDFNDY 1390
Qy 1591 GTQIPKEEKWSEKSEKPEKTAFFKXOTILSLNACESHAIAINEGKNKEIETWAKQGR 1650
Db 1391 ELYIPKQDOEAD-----FDGLL-----DHPE----- 1411
Qy 1651 TELCSQNPVLKRHQRETRITLQSDQBEIDYDDTISVEMKKEBFDIYD-EDENOSPS 1709
Db 1412 -----EYEXVEKQDYS---KTADVQALDATSQHLLKWA 1442
Qy 1710 FOKKTRHYTAAVERLDYGMSSPHVLRNRAQSGSVPOFKKVVQFEDTGDGFTQPLRYG 1769
Db 1443 GDANTYITISVEEENDYA-GVQRRLDKTAQNERPTVFRKVVFRYLDSTFIRDING 1501
Qy 1770 ELNEHGLGPGYTRAIVENIMVTFNQASRPYSFYSSLISY-----ESDQOGAEP 1821
Db 1502 EMDEHGLGILPLKAEVDQTMVVFPRNARSRPSYSLHANGVKYKQMEGLSYDDESPYWK 1561
Qy 1822 RKNFKVNETKTVFKVQHMAFTKDBPCAKAWAYESVDLEKDVHSGLIGPLLVCHTNT 1881
Db 1562 QDVAFPNGFTTMTWINTPKSGQNNESDCRTWYTSAVNFERDINSGLIGPLLVCKGT 1621
Qy 1882 LN--PAHGRQVTVQEFALFTTIEDETKSWYFTENMERNCRAPNIQMEDPTFKENYRFA 1939
Db 1622 LDKKPEDRR-----EFVLLFMTDENKSMLEYENRQRIERKNRVVX-DPNFQDLKFDA 1675
Qy 1940 INGYINDTLPCLWAQDQRIWLLSNGSNENTHSIHFSGHVTVKKEEYKALNLYP 1999
Db 1676 ING-ITYSLKGLRMYTNQLAKWHLINSGPKDLHSHVHEHQTQINKELKDHQGVPLLP 1734
Qy 2000 GVPEVTEMLPCKAGINRVECLIGEHLHAGMSTLF-----LVISNKKQ 2041
Db 1735 GGFATLEMLPSKGLQWLESEVGLSQORGMTLFLPDESYYKRVQCVLRSYLIYFAVCD 1794
Qy 2042 TPLGMSGHIRDFTQITASGOYQWAPKLABLHYSGSINAWST-KEPFSWIKVLLAPMII 2100
Db 1795 HPLGLISGTQDEQITASDRGQWYPLHSLHNTGKYNASTSSEPQYQLQVDFQRPVVI 1854
Qy 2101 HGIKTOAROKFSELYISQFIIMVSLDGKQWQYRGNSTGT-----LMVFF----- 2146
Db 1855 SKVATQAKQFLTHFNFLNTISYSTDKKWKIYKGDSDAVRKANIKHTNVFFNARQTFE 1914
Qy 2147 GNVDSGIGKINFPPIIARIYLRHPHYSIRSTFLRMELMCDLNSCMLPGVESKAISD 2206
Db 1915 GNAEAYETKENIFPPPLIGYVRLPLHPSYNTFVRLYYGCELDGCGSVFLGMEKGLIDD 1974
```

```
Qy 2207 AQITASSYFTINMFA-TWSPKARLELQGRSNARPOVNNPKEWLQVDFQKTMKVTGVTTQ 2265
Db 1975 SKITASSVASNWYSGQHPWYARLNKQCTANAWAKNNDIQPWIQVELKEVKKITGVTTQ 2034
Qy 2266 GVKLLISMTYKEFLISSQDCHOWTLFFQNG--KVKVFQGNQDSFTPVVNSLDPPLLTR 2323
Db 2035 GAKSGMGNFVRSYILEYSESGREWMKYTDDEDEQKLFQGNNTDNGQIKNIYIPIFSR 2094
Qy 2324 YLRTHPOSVWHQIARMEVLQCE 2346
Db 2095 FIRIIPKQWQSVTVRIELGCD 2117
RESULT 6
088783 PRELIMINARY; PRT; 2183 AA.
AC 088783;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Murine coagulation factor V.
GN F5.
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98282202; PubMed=9616155;
RA Yang T.H., Cui J., Rehmulla A., Yang A., Mousalli M., Kaufman R.J.,
RA Ginsburg D.;
RT "The structure and function of murine factor V and its inactivation by
RT protein C.";
RL Blood 91:4593-4599(1998).
CC -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
DR EMBL; U52925; AAC99553.1; -.
DR PIR; T42764; T42764.
DR RSP; P12259; 1CZT.
DR MGD; MGI:89382; F5.
DR GO; GO:0005507; P:copper ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FAS8C.
DR InterPro; IPR008979; Gal_Bird_like.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FAS8C_2; 2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00022; FAS8C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
SQ SEQUENCE 2183 AA; 247228 MW; BFOA8AA723F60317 CRC64;
```

Query Match 22.3%; Score 2771.5; DB 11; Length 2183;

Best Local Similarity 29.8%; Pred. No. 3e-156;

Matches 749; Conservative 409; Mismatches 851; Indels 503; Gaps 73;

```
Qy 8 CFPLCLL---RFC-----FSATRRYYLGVAVELSWDYMQSDGLGELPYDARFPPVPKX 56
Db 7 CFLLVVLGTRWAGWSHQAEAAQLQFYVAQGLIWNYPH-----PTD 51
Qy 57 FFPNTSVVYKTLFVEFTDHLFNIAKPRPPMGLGPTIOAEVDTVITLKNMASHFVS 116
Db 52 PSLSNIPSKVIYREY-EQYFKKEKPRSSNGLLGPTLYAEVGVIVKHFFNKADKPLS 110
Qy 117 LRHGVSYWKASEGABYDDQTSQREDDKVPFGSGHTVWQVLKENGPMASDPLCLTYS 176
Db 111 IHFQGIKYKFGSEGASAYADFTFAERKQDAVAPGEYTYEWISVSDSGTTPDPPCLTHI 170
Qy 177 YLHSDVLDKSLGILGALLVCREGSLAKEKTQL--HKFILLFAVDFDEGKSHSSTKNS 234
```


Db	983	-----KTTSVKTSGGHNWTVYDGTGTHK-----IYSAELPDDM	1011
Qy	1476	REWGLGTGATSVTVYKKVENVLPKPDLP-----KTSCKVELLPKVHIYQKDLFPPTET	1529
Db	1012	KXYFEM-SQTNKKTKRVN-----RPHPQKGHGKTKRREYKFO-----PRSGLPSP	1061
Qy	1530	SNQSGHLDLVSGSLLOGTEGAIKWNEANRP-GKVPFLRVATESSAKTSKLLDPLAWN	1588
Db	1062	RGFNP-----LMSPRG-----RPQGLQPVNN--DETJINMPVWIGVPRPDES	1102
Qy	1589	HYGTQIPKEWKSQEKSPKTAFTAKKCDTILSLNACESNHAIAMAINEGQKPLEIETWAKQ	1648
Db	1103	DYELYLPQGE-----	1112
Qy	1649	GRTERLCSQPPVLKRHQREIETRTTLLQSDOEEDYDDTISVEMKKEDFDIYEDENQSPR	1708
Db	1113	-----PDHLVDYQNVK-----ANEYEVYNYKDPYRSNEDAKNLHLHQRKYLENR	1158
Qy	1709	SFOKTRHYFIAAVERLWYGMSSSPHVLNRNAQSGS---VPOFKKVPQEFDTGSGFTOP	1768
Db	1159	--DKDVRTYFIAEEVQWDY---ACVGQRRRRPQGNIRHTKFIKVPFLYLDSSFRTP	1212
Qy	1766	LYRGELNEHLGLLPVIRAEVDENIMVTFRNQSRSPFSYSSLLISYEEDORQA---EPRK	1823
Db	1213	EVRGDEHLGLGPVKAQVGTIMVVFKNASRFFSLHPNGVYSYK-QTECLSYEDGS	1271
Qy	1824	NF-----VKPNETKTYFWKVQHHMAPTKDEPDCKAWAYFSDVLEKDVHSGLIGPLLV	1876
Db	1272	NYWKYDNEVOPGATTYIWNVNPWVGPTSDENCRTWAYYSGVNERDILHSLIGPLLV	1331
Qy	1877	CHTNTLNPAHGRQVT-VQEPALFPTTIDETKSYWFTENMERNCRAPCNIQMDPTKENV	1935
Db	1332	CQKGTLN---QELTNTREPMFLFMTFDESQSYFDRNRIMLRSGWR-KVMDPDIMENL	1386
Qy	1936	REHAINCYIMDTPLGLVMAQDQIRVLLWSGNSNENIHSHPSCHYFVTEKK--BEYKVAL	1994
Db	1387	KFHSINGIIYN-LKGLRMYTNQLVSWHLINNGSPKDFNSVHPHGQTFPHKKKTTSYQAV	1445
Qy	1995	YNYLPGVFETVEMLPKAGRWRYECLIGEHLHAGMSTFLVYGNKCTPPLGMAASHIRDF	2054
Db	1446	YPLPGSFATLEMYPSPGHWQLWETEYGTNOEKGMTFLVLADDCHPLGLSGSVKND	1505
Qy	2053	QITASGOYGOWAPKLARLHYSGSINAWSTKEPSPSWIKVDLLAPMIHGIKTQGARQKFS	2114
Db	1506	QITAINTRGYWEPLARLHNQCKYNAWSTDQNYSWIQVDQRPVVISQVATQAKQFFQA	1565
Qy	2115	LYTSQFIIMTLDGKKWQTGRNSTGLMYFPGNVDSGGLKNIENPPIIARVIRLHPH	2174
Db	1566	QYYSYIYSYSDNRSSYFKGSDRDDIKVFTGNNNPYDKKCTFFFPFLPLGRFIRFHPLK	1625
Qy	2175	YSRSTRMLBMLGMDLNSCSPGLMESKAIQITASSYFTNMFA-TWSPSKARLHLQ	2233
Db	1626	WYNKATLRMPFYCGELDGCSPVLMGESLIEDHQITASSASKWYSQWTFPLGRLNKEG	1685
Qy	2234	RSNAWRPQVNPKEWLQVDFQKTMKVTVGTITQGVKSLLTSMYKBEFLISSQDGHQWTLF	2293
Db	1686	TINAWQAKYNDMDQWLQVLTQVKKITGITQGAKFILGAEMFVTSFSLQYSHDGINWHPY	1745
Qy	2294	FQNGKV--KYPQGNQDSFTPVNSLDPLATRYIRIHPQSWHIOALRMEVLGCEAQ	2348
Db	1746	TDDGVPKAFIMGNTNNNDHYKMYPIFISRFVIRPKSGMIGSIPMRNEPLGDCDE	1802
RESULT 8			
Db	Q804X3	PRELIMINARY; PRT; 1377 AA.	
AC	Q804X3;	AC	
DT	01-JUN-2003 (TrEWBLrel. 24, Created)		
DT	01-JUN-2003 (TrEWBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TrEWBLrel. 25, Last annotation update)		
DE	Coagulation factor VIII precursor (fragment).		
GN	F8.		
OS	Gallus gallus (Chicken).		

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Davidson C.J., Hirt R.P., Lal K., Snell P., Elgar G.,
RA Tuddenham E.G.D., Mcvey J.H.;
RT "Comparative sequence analysis and molecular evolution of blood
RT coagulation genes from Gallus gallus and Fugu rubripes.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AP465272; AAO33367.1; -
DR GO; GO:005507; Fcopper: ion binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR000421; FA58 C.
DR InterPro; IPR008979; Gal_bind_like.
DR Pfam; PF00394; Cu-oxidase; 1.
DR Pfam; PF00754; F5_F8_type C; 2.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS50022; FA58C_3; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
FT NON TER 1 1
SQ SEQUENCE 1377 AA; 152728 MW; 15ACDF7F765665CD CRC64;

Query Match 20.2%; Score 2511.5; DB 13; Length 1377;
Best Local Similarity 36.3%; Pred. No. 5e-141;
Matches 618; Conservative 210; Mismatches 491; Indels 385; Gaps 48;

Qy 706 LWTIGCHNSDFNRGEMTALLKVSSCDKNTGDYYEDSYEDISAYLLKNNNAIEPRSPS--Q 763
Db 1 VVTILGCNVPDFRDRGMEAKFTVTCQCLE-GSLDEEYEDYE---BEDFELQPRGFSKAK 55

Qy 764 NSRHPSTROKQFNATIPENDIEK-----TDPFAHRTMPKIQNVY-SSDILMLLRQSP 817
Db 56 NKRRCPCVSEQFNNVIS--PKNGTRKPAICLTES--SHRALFHNITNPSPSGNSATSPFGSP 112

Qy 818 TPCHLSISDIQEAKEY-----TFSDDPSPGCAIDSNLSLSEMTHPRPQ.H 861
Db 113 HPDQVTSSTLPETTYDVPVSYESFLADEELSKTISQCGAGALPSK-----H 160

Qy 862 HSGDMVTPESGQLRLNEXLGTTAATELKKLPKVSSTNNLSTIPSNLAAGTDNTS 921
Db 161 ISGEVRGTVTSRGLYQSKAPEDAMVERK-----VTNVL 195

Qy 922 SLGPPSPMVHYDQLDTTLFGKKSSPLTESGGPLSISENNDSKLIES-GLMNSQSSWG 980
Db 196 EVQBEFA-----KAAVVQAGGTLQMLETFETQKPMATYDLMNWTFFAAG 238

Qy 981 KNVSSTESGLFKGKRAHGPAALLTKDVALPKVSIILKTKNTKSNNSATNKTHTIDGPSLL 1040
Db 239 K-----GPI-----CQTRSSFOQDGP--- 254

Qy 1041 IENSPSPVQWNILESDDTFEKVYTP-ITHDRMLMDKNAFALRLNFKSNKTTSSKNMVMVQK 1099
Db 255 -EHS---LGIQVTSSEGADVTPPLNLHESREPSNTGPTLGSNHSSTTDS----- 300

Qy 1100 KEGIPIDPAQNPMSPFMFLFLESARWIIORTGKNSLNSGQSPSPKQLVSLGPEKVEG 1159
Db 301 ---PLGPSARTEDIG-----TSQSHSVESNRSSSELDAKLRPKHVYSQGFNSLGG 350

Qy 1160 QNFUSEKNVVGKERTKDVGLKEMVFPSSRNLFJTNLNLHNNHTNCKEIKIQRB--- 1216
Db 351 ENI-----SFSDLCRVOEQILTMKNSLPLANSMEAKGTFAHNSNLLERSRYL 398

Qy 1217 -IEKKETLIIQENVVLPQIHTVTGTFKFMKNLFLILSTRQNVGEGYDGAYPVLQDFRSLND 1275
Db 399 PTERDELILLEAV---PQDATATKQLPETDLSVLQPSNL----- 434

Qy 1276 STNFTKXHTAHFSKKGBEENLEGLNGQTKQIIVEKYAKTTRISPNVTSQNFVTRSKALK 1335

Matches	643;	Conservative	379;	Mismatches	779;	Indels	721;	Gaps	80;
QY	86	PWGHLGPTIQAEVYDVTIVITLKWMASHPVSLHVGVS-							124
Db	46	PSLGLGLPTLYAEVGDITIKVHPRNKADKFLSIHPQGI	RYSKPSE	QAQLPAEBEMDFAA	PF	I	105		
QY	125								124
Db	106	LRLCLDFPESLSQOALVLLPKRKTGFVTP2LLE	GVGVHRLSMGLYTNKRC	GI	FTLGGTG	165			
QY	125								139
Db	166	FMCPRSELHAGACGDRGGERPERASAPFLHTGE	WNKTLDFVAVTS	PGASYLCHTSP	225				
QY	140	REXEDDKVPPGGSHTVVQVLKANGPMASDPCLTY	SYLSYLVLDLKDLSGLIGALLVCR	199					
Db	226	AREKDDAAGAEEYTYEWFIGDSGPTDPDPCLTH	IYYSYENLAQDRN	SLGILGLLICK	285				
QY	200	EGSLAKEKQTLL--HKFILLFAVDFBGKSWHSE	TNLSLMQDRDAASARA	WPKHVTWGVV	257				
Db	286	EGVLTGDTQKMPDKGHVLLFAVDFDESKW--							329
QY	258	NRSLPGLICHRKXSVYVHVMGTTPEVASIFLE	GHTFLVRNHR	QASLE-	309				
Db	330	NRTMPDVTVCAYDHVSWHLIGMSGPELSIHP	QGVLEQNQHKTSTVLLQ	TGVGIF--RP	389				
QY	310	ITFLTAQTLMDLQFLLSCHISSHQHDGWEAVK	YVDSCPESPQLIMKNNE	SAEDYDDDL	369				
Db	390	IAANTVGQYIASRGDLVLPWA?--ZAAGQ	AVIDIKCPK--		429				
QY	370	TDSEMDVVRDDDNSPSFIQIRSAVKHKPTV	HYIAAEEEDMDYAPL	VLA--PDORSYKS	428				
Db	430	TRASKTLTR-----EQRHHKRWEPF	IAEBEWINYAPVIPAN	KDKHSYEL	475				
QY	429	OYLNGEQRIGRYKKVRPMAYTDEFTKREAT	QHESGILGPLLYGEVGD	TLLIIFKQ	488				
Db	476	SPGMWPFDDSNRN--NGSRMLATVPC	RLRUSHQFYKLGA	L-----	518				
QY	489	SRPVNTYEHGITDVRPLYSRRLPKGVKHL	KDFPLPGEI	PKYKWTVTVEDGPTKSDPRCL	548				
Db	519								533
QY	549	TRYYSFVNMRDLASGLIGLICVYESVDQ	KGNQIMSKRNILP	SPVDFDNRSWLT	608				
Db	534	VRHMPFLR-----VADMEQVVFAMF	DNKSWI	ED	565				
QY	609	NIQRFLPNPAGVQLEDEPFOASIMHSING	VFQSLQ--LSVCLHEVAYVILS	IGAQTDF	667				
Db	566	NINKFCENPEVKKDDPKFTESNMINTING	VYVPESTLGFCD	DDAVQVHFCVSGTHDDI	625				
QY	668	LSVPFGYTFKHKKVYEDTITLFPFGSET	VFMGMEPNGLMI	LOCHNSDFNRGWTALLKV	727				
Db	626	LTVHFTGHSFYGRHEDTTLFPM	SGESVTVMNDNIGTWMLTMSN	PRRRLRFRD	685				
QY	728	SSCDKNTGYVEDSYEDI	SAYLLSKNNAIPRFSQNSR	HPSTRQFNATITP	END--I	785			
Db	686	VKCNRDD--DEDSYE--IY-----	QPLSETMTTKIHDSV-----		ENDFGI	725			
QY	786	EKTDPEFAHRTMPMKIQNVSSDLMLLRQS--	--PTPHGLSLDQAEK	YETESDDPSPG	842				
Db	726	ENEDDDYQY-----ELASTLGIR	FRKELLNPKEDBNL	TALALENSSEFISL	STDG	777			
QY	843	AIDNNNS--LSEWTHFRPQLHSGDWV	TFPESGLQRLNEKLT	TAATKLLDKFV	VSST	900			
Db	778	VVDSKSENLSKIT-----	NNLNDSQRT-----	LSC	804				
QY	901	SNNLITIPSONLAAGTDNTSSLGPP	SPMVHYDSQDLTTLFGK	SS-----	PLATBSGG	953			
Db	805	SGATAGILLNL--TGLGNVILN--PSTE	VHSSSYENDMED	DPQSNITVVIPLG	AKG--	861			
QY	954	PLSLSEENNDSKLLFESGLMNSCES--	SWGNKVSSTESGR	LFKGRHGPALLTK	ONALFX	1011			
Db	862	--SSGRBOTKPTIKTGPRHAKRFSNMK--							892

[illegible]

1837 KVQHMAPTKDEDFCKAWAFSDVLEKQVHSGSLGILLVCHTNTLNPAGRQVTVQFEA 1896
235 YANRSPLOFGACRACRQVYSDNLEKQTHSGLIGLILCEKGTFSASDNRPINTDFF 294
1997 LFTTIDETKSWFTENMRNCRAPCNIQMEDPTFKENYRPHALNGVIXDTPLGVLMAQD 1956
295 LLEWVFEERSWYF---DKRSRKPCTEKTQE---MOQCHKFYALNG-IUNNLOGLRMVEG 347
1957 QRTRWYLLSGSNEHISIHFSGHVFTVRKEEYKMAIYNLYPGVEETVEMLPKAGIWR 2016
348 ELVKNHLLNMGVPKDIIHVHFGHTFIEQLEAKYQJGTYYLLPGSFQIVENKMPQRPQWL 407
2017 VECUIGELHAGMTLPLVATSNKQOTPLGMASGHIRDFQITASQYQGWAPKLARLYSG 2076
408 LDTVEGEVQQAQASVYLVLEKGRCPMGASGVLDSQINASHHIDYWEPLARLNSG 467
2077 SINAMST--KEPPSWIKVLLAPMIHIGIKTOGARQKFPSSLYISQFIMYSLDKKKQTY 2134
468 TYNAMSTMTTDELPIWQVDFORQVLLTGIQOGAKHFLKSFYQKLFIVYSTDKRKNWF 527
2135 RGNSTGTLWVFGNVDSGKIHNFNPPIIARIYRLRPHYSIRSLRMELMGCDLNSCS 2194
528 RGDSPARKIEGNSDAYGIKENIDPIIARIYRVYPKQAYNEPTLRMELMGCEVDACS 587
2195 MPLGMESKAISDAQITASSYFTNFMATWSPSKARHLQCRSNAMRPQVNPKEWLQVDFQ 2254
588 LPLGMENGEIKNTQITASSVTSWFTWEPSLARINQKGINAWRAENNNQOWLQIDLL 647
2255 KTMKVTGVTGQVSLTSMVKVFLSISSODGHWTLFFQNGK--VKVFGNODSFTPV 2312
648 TIKKITALATQGVSEVTENFVKTYVILYSNQGSEWKSFTEDSSSVAKVFGNEDSRGHV 707
2313 VNSLDPLLTLYRLRHPOSVEQALRMEVLGCE 2346
708 KHFNPPILSFIRVPTWNGIALRVELFGCD 741
RESULT 13
Q80Y80
ID Q80Y80 PRELIMINARY; PRT; 1156 AA.
AC Q80Y80;
DT 01-JUN-2003 (TRENBL-rel. 24, Created)
ET 01-JUN-2003 (TRENBL-rel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBL-rel. 25, Last annotation update)
DE Hephaestin (Heph protein).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins E.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Suecow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Kulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]
RN SEQUENCE FROM N.A.
RP STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RL Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC048237; AH48237.1; -
DR EMBL; BC054442; AH54442.1; -
DR GO; GO:0005507; F: copper ion binding; IEA.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR008972; Cupredoxin.
DR InterPro; IPR002355; Cu ox copper_BS.
DR Pfam; PF00394; Cu-oxidase_3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1156 AA; 129567 MW; 5F110333599DC1874 CRC64;
Query Match 11.2%; Score 1390; DB 11; Length 1156;
Best Local Similarity 20.0%; Pred. No. 4.3e-74;
Matches 412; Conservative 180; Mismatches 413; Indels 1052; Gaps 26;
QY 20 ATRYKILGAVELSWYXQSDLGELPVDARPPRPVKSF-----PNTSVVYKTLFVETD 75
DB 24 AIRYVYLGIDMQWYAPKGRNVITNLTNDTVASSFLSKGNRIGSSYKTKYKEYSD 83
QY 76 HLFNIAKPRPMGLLPTQAEVYDVTWITLKNWASHVSLHAGVSYWKASGEAYDD 135
DB 84 GTYTEIAKPAWLGFLGLQAEVGDVILLHLKNFASRPYTIHPHGVFYEKDSGSLYPD 143
QY 136 QTSOREKDDKVPFGSGHTYVQVLKENGPMASDPLCLTYSLSHVDLVKDLNLSGLTAL 195
DB 144 GSSGLYKADDSVPPGGSHVYNWSIPESHAPTEADPACLTWYHSHVDAPRIATGLIGPL 203
QY 196 LVCREGSL---AXEKTQTLHKFILLFAPVDPDEGKSWHE-----TKNSLMQDRDAASAR 245
DB 204 ITCKRGTLDGNSPPQRKVDVHNFFLFPVIDENLSWLDNDNIATYCSDPASVDKEDGAFQ 263
QY 246 AWPKNHIVNGVYANSLPGLICGCHKSVYVHVGITTPPEVHSIFLEGHTFLVRNERQASL 305
DB 264 DSNRMHAINGEVFNGLPELNSCAOKRVAMHLFGMNGEIDVHTAFHQMGLSIRGHTDVA 323
QY 306 EISPTITLTAQTLMLDGLQFLLSCHISSHQHDGMEAYKYVDSCPEEPQLIMKNNERAEDY 365
DB 324 NIFPATFVTAEMVFKSGSTWLISCEVNSHLRSGMQAFYKVDSCSMDPPV----- 372
QY 366 DDDLTDSEMDVFRDDNSPSFIQIRSVAKKPKTWHTYIAAEEDWDYAPLVAPDPRS 425
DB 373 -DQUTG-----KVRQFIQAHETQMDYGP--IGYDGT 402
QY 426 YKS-----QYLNNGPQRIGKRYKVRWYATDETFFKTREATIQUES--GILGPLLYG 474
DB 403 CKSLREPOSGDPKVFQKSSSSIGGTYKVRVEAQDETFOBRVHQEETHLIGLGPVIRA 462
QY 475 EVGDTLLIIFKNQSRPNYVPHGIDVRLPYSLRRLPKGVKH---LKDFFLIPGFIKYK 531
DB 463 EVGDTIQVVFYNRASQPSFIQPHGVFYEKNGEGTVNDGTSHPKVAKSF-----EKVTYY 517
QY 532 WTVTVEDGFTKSDERCLTRYSSVFNMERDLASGLILICYKESVDQKGNQIMSKRN 591
DB 518 WTVVPFHAGTQDFACLTWMTVFSAADPFRDNSGLVGLLVCKAGALGADGQKGVKKEF 577
QY 592 VILPSVFDENSWYLTENIQRFLENPAG-----VOLEDPE--FQASNIMHESINGVYFDSL- 644
DB 578 FLFTVTFDENSWYNSAN-----QAAGMLDSRLSEDEVEGFQDSNRMAHNGFLSNLP 631
QY 645 QLSVCLHEVAYWYILSIGAQDTFLSVFFSGYTFKHQWYEDTLTLPFSGTGFVMSMNP 704
DB 632 RLDCKGDTVAHLLGLGTETDVHGVMEGNTVQLQGRKGAVMFLPHTFTVTAIMPQDNP 691

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:01:28 ; Search time 22.5 Seconds
(without alignments)
5394.340 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQIELSCFFLCRLRPFCSA.....WFOIALMEVLGCEAQDLY 2351

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*

2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*

3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*

4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*

5: /cgn2_6/ptodata/2/iaa/PCRTUS COMB.pcp.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12399	99.9	2351	4	US-10-133-907-4
2	12393	99.8	2351	6	Sequence 4, Appli Patent No. 5171844
3	12390	99.8	2351	1	US-08-366-851A-2
4	12386	99.8	2351	1	US-08-121-202-2
5	12377	99.7	2351	6	Sequence 2, Appli Patent No. 5422260
6	12382	98.9	2332	1	US-07-864-004B-4
7	12282	98.9	2332	1	Sequence 4, Appli Patent No. 5422260
8	12282	98.9	2332	1	US-08-251-937A-4
9	12282	98.9	2332	1	US-08-212-133A-2
10	12282	98.9	2332	1	US-08-474-503-2
11	12282	98.9	2332	3	US-08-670-707A-2
12	12282	98.9	2332	3	US-09-037-601-2
13	12282	98.9	2332	4	US-09-315-179-2
14	12282	98.9	2332	4	US-09-523-656-2
15	12282	98.9	2332	5	US-09-523-656-2
16	12275	98.9	2332	3	US-08-276-594A-2
17	12232	98.5	2332	3	US-09-324-867-3
18	9474	76.3	2343	3	US-09-324-867-2
19	8820	71.0	2319	1	US-08-212-133A-3
20	8820	71.0	2319	1	US-08-474-503-6
21	8820	71.0	2319	2	US-08-670-707A-6
22	8820	71.0	2319	3	US-09-037-601-6
23	8820	71.0	2319	4	US-09-315-179-6
24	8820	71.0	2319	4	US-09-523-656-28
25	8820	71.0	2319	5	US-09-523-656-28
26	8715	70.2	2304	3	US-09-324-867-4
27	8338	67.2	1661	2	US-08-882-083-2

28	8338	67.2	1661	2	US-08-558-107-2	Sequence 2, Appli
29	8338	67.2	1661	2	US-09-243-539-2	Sequence 2, Appli
30	8057	64.9	2133	3	US-08-670-707A-37	Sequence 37, Appli
31	8057	64.9	2133	3	US-09-037-601-37	Sequence 37, Appli
32	8057	64.9	2133	4	US-09-315-179-37	Sequence 30, Appli
33	8051	64.8	2133	4	US-09-523-656-30	Sequence 30, Appli
34	7940.5	64.0	2115	3	US-09-324-867-5	Sequence 5, Appli
35	7381	59.5	1471	1	US-08-683-839B-3	Sequence 3, Appli
36	7208	58.1	1438	4	US-09-209-916-1	Sequence 1, Appli
37	6205	50.0	1467	4	US-09-523-656-38	Sequence 38, Appli
38	6103	49.2	1443	2	US-08-670-707A-39	Sequence 39, Appli
39	6103	49.2	1443	3	US-09-037-601-39	Sequence 39, Appli
40	6103	49.2	1443	3	US-09-315-179-39	Sequence 39, Appli
41	2771.5	22.3	2183	3	US-08-746-111-5	Sequence 5, Appli
42	2425.5	19.5	541	1	US-08-121-202-4	Sequence 4, Appli
43	2258	18.2	439	3	US-08-448-722A-5	Sequence 5, Appli
44	2258	18.2	439	3	US-08-189-309B-5	Sequence 5, Appli
45	2190.5	17.6	868	1	US-07-864-004B-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-133-907-4
; Sequence 4, Application US/10133907
; Patent No. 667369
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; FILE REFERENCE: 6627-P1170
; CURRENT APPLICATION NUMBER: US/10/133,907
; CURRENT FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-907-4

Query Match	98.98	Score	12399	DB	4	Length	2351
Best Local Similarity	99.98	Pred. No.	0				
Matches	2348	Conservative	1	Mismatches	2	Indels	0
Gaps	0						
Qy	1	MQIELSCFFLCRLRPFCSA	TRRYVLGAVELSDWYMQSDLGELPVDARFPPRVPKSPFN	60			
Db	1	MQIELSCFFLCRLRPFCSA	TRRYVLGAVELSDWYMQSDLGELPVDARFPPRVPKSPFN	60			
Qy	61	TSVYKKTFLVEFTDHLFNIAKRP	PPWMLGPTIQAEVDTVVITLKNASHVPVSHAV	120			
Db	61	TSVYKKTFLVEFTDHLFNIAKRP	PPWMLGPTIQAEVDTVVITLKNASHVPVSHAV	120			
Qy	121	GVSWKASEGAEXDDOTSOREKED	KVPFGSHYVWQVLKENGPMASDPLCLTYSLSH	180			
Db	121	GVSWKASEGAEXDDOTSOREKED	KVPFGSHYVWQVLKENGPMASDPLCLTYSLSH	180			
Qy	181	VDLVKDLNSGLIGALLVCREGSL	AKETQTLHKFILLFAVDEGKSWHSEKNSLMQDRD	240			
Db	181	VDLVKDLNSGLIGALLVCREGSL	AKETQTLHKFILLFAVDEGKSWHSEKNSLMQDRD	240			
Qy	241	AASARAPKQHTVNGYVNRSLPG	ILGCHRSVYWHVIGMGTTPVHIFLEGHTFLVRNH	300			
Db	241	AASARAPKQHTVNGYVNRSLPG	ILGCHRSVYWHVIGMGTTPVHIFLEGHTFLVRNH	300			
Qy	301	RQASLEISPTIFLTAOTLLMDL	QFLSCHISSHODGMEAYKVVDSCPEEPQIMKNE	360			
Db	301	RQASLEISPTIFLTAOTLLMDL	QFLSCHISSHODGMEAYKVVDSCPEEPQIMKNE	360			

361 EAEDYDDDLTSEMDVAFDDDNPSFQIORSVAKKHFKTWHVYIAAEEEDWDVAPLV.A 420
 361 EABYDDDLTSEMDVAFDDDNPSFQIORSVAKKHFKTWHVYIAAEEEDWDVAPLV.A 420
 421 PDDRSYKQYLNNGPQRIGRKYKVRWAYTDFKTREAQHESGILGPLLYGEVGTLL 480
 421 PDDRSYKQYLNNGPQRIGRKYKVRWAYTDFKTREAQHESGILGPLLYGEVGTLL 480
 481 LIIFKQASRYNYPHGITDVRPLYSRRLPKGVKHLKDPILPGEILFKYKWTVTVEDGP 540
 481 LIIFKQASRYNYPHGITDVRPLYSRRLPKGVKHLKDPILPGEILFKYKWTVTVEDGP 540
 541 TKSDPRCLTRYSSFWMERDLASGLIGPLLI CYKESVDQGNQIMSKRNVILFSVFDE 600
 541 TKSDPRCLTRYSSFWMERDLASGLIGPLLI CYKESVDQGNQIMSKRNVILFSVFDE 600
 601 NRSWILTNIORFLNPAGVOLDEDFEASNTMHINGYVFDLSQLSVCLHEVAVWYILS 660
 601 NRSWILTNIORFLNPAGVOLDEDFEASNTMHINGYVFDLSQLSVCLHEVAVWYILS 660
 661 ICAQTDFLSVFSGYTFKHKMYREDTLLFPFSGETVFSMENPGLWILGCHNSDFNRG 720
 661 ICAQTDFLSVFSGYTFKHKMYREDTLLFPFSGETVFSMENPGLWILGCHNSDFNRG 720
 721 MTALLKVSCKNTGDYVEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI 780
 721 MTALLKVSCKNTGDYVEDSYEDISAYLLSKNNAIEPRSFQNSRHPSTRQKQFNATTI 780
 781 PENDIEKTDPAHPTMPKTONVSSDLMLLRQSPTHGLSLSDLOEAKYEFSDDPSS 840
 781 PENDIEKTDPAHPTMPKTONVSSDLMLLRQSPTHGLSLSDLOEAKYEFSDDPSS 840
 841 PGALDSNNLSMTHFRQLHSHGDMVTPESGLQLRLNEKLGTTAAATELKLDFKVSST 900
 841 PGALDSNNLSMTHFRQLHSHGDMVTPESGLQLRLNEKLGTTAAATELKLDFKVSST 900
 901 SNNLSTIPSDNLAAGTNTSSLGPPSMFVHYDSQDITLFGKXSSPLTESGGPLSLSEE 960
 901 SNNLSTIPSDNLAAGTNTSSLGPPSMFVHYDSQDITLFGKXSSPLTESGGPLSLSEE 960
 961 NNDKLLSGLKNSQESGKNVSTESGRIFKGRAGPALLTKDNALFKVSISSLKTN 1020
 961 NNDKLLSGLKNSQESGKNVSTESGRIFKGRAGPALLTKDNALFKVSISSLKTN 1020
 1021 KTSNNSATNRKTHIDQPSLLIENSFSVWQNILESUTEFKKVTPLIHDRMLMDKNATLRL 1080
 1021 KTSNNSATNRKTHIDQPSLLIENSFSVWQNILESUTEFKKVTPLIHDRMLMDKNATLRL 1080
 1081 NMSNKTSSKNMVMVQKKKEGPIPDQONPDMSPFFKMLFPESARWQRTHGKNSLNSG 1140
 1081 NMSNKTSSKNMVMVQKKKEGPIPDQONPDMSPFFKMLFPESARWQRTHGKNSLNSG 1140
 1141 QGPSKQVLVSLGPEKSVGQNFLESKNKVVGKGEFTKDVGLKEMVFPSSRNLFUTNLDN 1200
 1141 QGPSKQVLVSLGPEKSVGQNFLESKNKVVGKGEFTKDVGLKEMVFPSSRNLFUTNLDN 1200
 1201 LHENNTNHOEKIIOBEIEKKEKTELIOENVVLPQHTVGTCTKPFKXNLFLLSTQNVEGSD 1260
 1201 LHENNTNHOEKIIOBEIEKKEKTELIOENVVLPQHTVGTCTKPFKXNLFLLSTQNVEGSD 1260
 1261 GAYAPVLQDFRSLNDSNTRKXHTAHFSGKGBEENLEGNGTKQIVERYACTTRISNPT 1320
 1261 GAYAPVLQDFRSLNDSNTRKXHTAHFSGKGBEENLEGNGTKQIVERYACTTRISNPT 1320
 1321 SOQNFTQSRKRALQFRLPLETELEKEIIIVDDTSTOWSKNMKHLTPSTLQIDYNEKE 1380
 1321 SOQNFTQSRKRALQFRLPLETELEKEIIIVDDTSTOWSKNMKHLTPSTLQIDYNEKE 1380
 1381 KGAIQOSPLSDCCLTRSHSIFQANRGPPLPIAKVSSPPSIRPIYLRVLFQDNSSHLPAAZY 1440
 1381 KGAIQOSPLSDCCLTRSHSIFQANRGPPLPIAKVSSPPSIRPIYLRVLFQDNSSHLPAAZY 1440
 1441 RKKDSGVQESSHFLQGAKKNNLSLAILTLEMTQDQREVSGSLGFSATNSVTYKKVENTVLP 1500

1441 RKKDSGVQESSHFLQGAKKNNLSLAILTLEMTQDQREVSGSLGFSATNSVTYKKVENTVLP 1500
 1501 KPDLPKTSIGKVELLPKHVHYQKOLPPTETSNNGSPGHLDVEGSLLOQTEGAIKWNEANRP 1560
 1501 KPDLPKTSIGKVELLPKHVHYQKOLPPTETSNNGSPGHLDVEGSLLOQTEGAIKWNEANRP 1560
 1561 GKVFFLAVATESSAKTPSKLLDPLAWNHYGTQIPKBEKMSQEKSPKTAPEKKDXTILSL 1620
 1561 GKVFFLAVATESSAKTPSKLLDPLAWNHYGTQIPKBEKMSQEKSPKTAPEKKDXTILSL 1620
 1621 NACESNEHAIANEGQNKPEIEVTWAKQGRTERLCSQNPVVLKHXHQBREITRTTILQSDQEE 1680
 1621 NACESNEHAIANEGQNKPEIEVTWAKQGRTERLCSQNPVVLKHXHQBREITRTTILQSDQEE 1680
 1681 IDYDDTISVEMKK3DFDIYDEDENQSPRSPOKKTTRHYFLAAVERLWDYGMSSSHVLNRR 1740
 1681 IDYDDTISVEMKK3DFDIYDEDENQSPRSPOKKTTRHYFLAAVERLWDYGMSSSHVLNRR 1740
 1741 ACSGSVPQFKVWFQEFDTGDSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASR 1800
 1741 ACSGSVPQFKVWFQEFDTGDSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVTFRNOASR 1800
 1801 PYSFYSSLIISYEEDORQAGAPRKNFVKPNETKTYFWKVQHMAPTKDFDCAWAYFSDV 1860
 1801 PYSFYSSLIISYEEDORQAGAPRKNFVKPNETKTYFWKVQHMAPTKDFDCAWAYFSDV 1860
 1861 DLEKDVHSGILGPLLVCHTNTLNPAHGRONTVOEFALPFTTFDETKSWYFTEENVERCRA 1920
 1861 DLEKDVHSGILGPLLVCHTNTLNPAHGRONTVOEFALPFTTFDETKSWYFTEENVERCRA 1920
 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQDQIRWYLLSMGSNENIHSIFSGH 1980
 1921 PCNIQMEDPTFKENYRPHAINGYIMDTLPGIWMADQDQIRWYLLSMGSNENIHSIFSGH 1980
 1981 VFTVRKKEEYKMAIYNLYPGVFETVEMLPKAGIWRVBCCLIGEHLHAGMSTLFLVYNNKC 2040
 1981 VFTVRKKEEYKMAIYNLYPGVFETVEMLPKAGIWRVBCCLIGEHLHAGMSTLFLVYNNKC 2040
 2041 QTPLOMASGHIRDQITASGQYGOWAPKARLHSHSGSINAWSTKEPFSWIKVDLLAPMII 2100
 2041 QTPLOMASGHIRDQITASGQYGOWAPKARLHSHSGSINAWSTKEPFSWIKVDLLAPMII 2100
 2101 HGIKTQGARQKFSLLIYSQFIIMTSLDGKKWQTVRGNSTGTLMVFFGNVDSGSGIKHIFN 2160
 2101 HGIKTQGARQKFSLLIYSQFIIMTSLDGKKWQTVRGNSTGTLMVFFGNVDSGSGIKHIFN 2160
 2161 PPIIARYIRLPHTHYSIRSTRMLMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMFA 2220
 2161 PPIIARYIRLPHTHYSIRSTRMLMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMFA 2220
 2221 TWSPSKARLHLOGRSNAWRPOVNNPKEWLQVDFOKTKVGTGVTQGVKSLITSMYVKEFL 2280
 2221 TWSPSKARLHLOGRSNAWRPOVNNPKEWLQVDFOKTKVGTGVTQGVKSLITSMYVKEFL 2280
 2281 ISSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
 2281 ISSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRIHPQSWVHQIALRM 2340
 2341 EVLGCQAQDLY 2351
 2341 EVLGCQAQDLY 2351

RESULT 2

5171844-2

; Patent No. 5171844

; APPLICANT: VAN COYEN, ALBERT J.J.; PANNEKOEK, HANS; VERBEET,
 ; MARTINUS P.; VAN LEEN, ROBERT W.

; TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY
 ; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS
 ; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM
 ; NUMBER OF SEQUENCES: 12

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/205.226
; FILING DATE: 10-JUN-1988
; SEQ ID NO:2
; LENGTH: 2351
5171844-2

Query Match      99.8%; Score 12393; DB 6; Length 2351;
Best local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MQIELSCFFLCLLRFCSATRRYYLGAVELSWDMQSDLGELVDARFFPRVPEKSPFN 60
DB 1 MQIELSCFFLCLLRFCSATRRYYLGAVELSWDMQSDLGELVDARFFPRVPEKSPFN 60
QY 61 TSVVYKXKTLFVEFTDHLFNIAKPPPPWGLLGP--TQAEVYDVTVVITLKNMASHVPSLEAV 120
DB 61 TSVVYKXKTLFVEFTDHLFNIAKPPPPWGLLGP--TQAEVYDVTVVITLKNMASHVPSLEAV 120
QY 121 GVSWKASGAEYDCQTSOREKEDDKVPFGSGSHYYVQVILKENGPMASDPLCLTYSYLSH 180
DB 121 GVSWKASGAEYDCQTSOREKEDDKVPFGSGSHYYVQVILKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDLSGLIGALLVCREGSLAKEKQTOLHKEKILLFAVDEGKSWHSETKNSLMQDRD 240
DB 181 VDLVKDLSGLIGALLVCREGSLAKEKQTOLHKEKILLFAVDEGKSWHSETKNSLMQDRD 240
QY 241 AASARAWPKMHTVNGVYVNRSLPGJIGCHRXKSVYWHVIGMGTTPVHSTFLEGTFLVRNH 300
DB 241 AASARAWPKMHTVNGVYVNRSLPGJIGCHRXKSVYWHVIGMGTTPVHSTFLEGTFLVRNH 300
QY 301 ROASLEISPIFLTAQTLMDLGOPLLSCHISSHOHDGMEAYVVDSCPEEPQLRMKNE 360
DB 301 ROASLEISPIFLTAQTLMDLGOPLLSCHISSHOHDGMEAYVVDSCPEEPQLRMKNE 360
QY 361 EAEDYDDLTLSEMDVVRFDNDSPFIQIRSAVAKKHPKTVWHYIAAEEEDWDVAPLVLA 420
DB 361 EAEDYDDLTLSEMDVVRFDNDSPFIQIRSAVAKKHPKTVWHYIAAEEEDWDVAPLVLA 420
QY 421 PDARSYKQYLNNGPQRIGRYKKVVRPMAYTDETPKTREAOHESGILGPLLYGVEGDTL 480
DB 421 PDARSYKQYLNNGPQRIGRYKKVVRPMAYTDETPKTREAOHESGILGPLLYGVEGDTL 480
QY 481 LIIFKNQASRPVNIYPHGITDVRPIYSRRLPKGVKHLKDFPLPGEIFYKWTVTVEDGP 540
DB 481 LIIFKNQASRPVNIYPHGITDVRPIYSRRLPKGVKHLKDFPLPGEIFYKWTVTVEDGP 540
QY 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQKGNQMSDKRNVLFSVFE 600
DB 541 TKSDPRCLTRYSSFVNMRDLASGLIGPLLYCYKESVDQKGNQMSDKRNVLFSVFE 600
QY 601 NRSWYLTENIQRFLENPAGVQLEDPEQASNIMHSINGYVFDLSQLSVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRFLENPAGVQLEDPEQASNIMHSINGYVFDLSQLSVCLHEVAYWYILS 660
QY 661 IGAQTDLSVPSFGVTEKHMYVEDTLTFPFSGETVFMSENPGILWLGCHNSFRNRG 720
DB 661 IGAQTDLSVPSFGVTEKHMYVEDTLTFPFSGETVFMSENPGILWLGCHNSFRNRG 720
QY 721 MTALLKVSCKNTGDYVDSYEDISAYILSKXNAIEPRFSQNSRHPSTROKQFNATTI 780
DB 721 MTALLKVSCKNTGDYVDSYEDISAYILSKXNAIEPRFSQNSRHPSTROKQFNATTI 780
QY 841 PCGAIIDNSNLSMTFRPOLHSGDMVFTPEESGLQRLNEKLGTGTAATSEKLLDKVSGST 900
DB 841 PCGAIIDNSNLSMTFRPOLHSGDMVFTPEESGLQRLNEKLGTGTAATSEKLLDKVSGST 900
QY 901 SNNLSTIPSDNLAAGTNTSSLGPPSPMPVHYDQSDLTTLFGKKSPLTESGPGLSLSEE 960

```

```

DB 901 SNNLSTIPSDNLAAGTNTSSLGPPSPMPVHYDQSDLTTLFGKKSPLTESGPGLSLSEE 960
QY 961 NNDSKLLESGIAMSQSSGKGNVSSTESGRLLFGKGAHGPAALLTKDNALPKVLSILLKTN 1020
DB 961 NNDSKLLESGIAMSQSSGKGNVSSTESGRLLFGKGAHGPAALLTKDNALPKVLSILLKTN 1020
QY 1021 KTSNNSATNRKTHIDGSPSLIENSPSVWQNLLESDETFKVKVPLIHDRMLMDKNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGSPSLIENSPSVWQNLLESDETFKVKVPLIHDRMLMDKNATLRL 1080
QY 1081 NHMSNKTSSKNMEMVQCKEGEPIPPDAQPDMSPFFKMLFLPESARWIQTHGKNSLNSG 1140
DB 1081 NHMSNKTSSKNMEMVQCKEGEPIPPDAQPDMSPFFKMLFLPESARWIQTHGKNSLNSG 1140
QY 1141 QGSPKQIVSLGPEKSVGEGQNFLEKKNVVGGEPTKQVGLKEMVPPSRNLFITLND 1200
DB 1141 QGSPKQIVSLGPEKSVGEGQNFLEKKNVVGGEPTKQVGLKEMVPPSRNLFITLND 1200
QY 1201 LHENNTNQEKKIQEBIEKKEETLIQENNVLPQIHTVTGTFKNFMKNLFLSTRQNVESYD 1260
DB 1201 LHENNTNQEKKIQEBIEKKEETLIQENNVLPQIHTVTGTFKNFMKNLFLSTRQNVESYD 1260
QY 1261 GAYAPVLQDFRSLNDSNTRTKKHTAHFSPKGBEENBGLGNQTKQIVEKYACTTRISPT 1320
DB 1261 GAYAPVLQDFRSLNDSNTRTKKHTAHFSPKGBEENBGLGNQTKQIVEKYACTTRISPT 1320
QY 1321 SQQNFVQSRKRALKQRLPLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKE 1380
DB 1321 SQQNFVQSRKRALKQRLPLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKE 1380
QY 1381 KGALTQSPSLDCLTRSHSIPOANRSPPIAKVSPFSIRPIYTRVLFDQNSSHELPAASY 1440
DB 1381 KGALTQSPSLDCLTRSHSIPOANRSPPIAKVSPFSIRPIYTRVLFDQNSSHELPAASY 1440
QY 1441 RKXDSGVQESSEHFQGAKKNNLSAILTLEMTGQREVSGISGTSATNSVTVKVENTVLP 1500
DB 1441 RKXDSGVQESSEHFQGAKKNNLSAILTLEMTGQREVSGISGTSATNSVTVKVENTVLP 1500
QY 1501 KFDLPKTSKGVKVELLPKHVYQKDLFPPTETSNQSGHLDLVEGSLLOGTEGAIKNENARP 1560
DB 1501 KFDLPKTSKGVKVELLPKHVYQKDLFPPTETSNQSGHLDLVEGSLLOGTEGAIKNENARP 1560
QY 1561 GKVPFLRVATESSEAKTSPKLLDPLAWDNHYGTQPKBEWKSQESPEKTAFFKKDTILSL 1620
DB 1561 GKVPFLRVATESSEAKTSPKLLDPLAWDNHYGTQPKBEWKSQESPEKTAFFKKDTILSL 1620
QY 1621 NACSNHAIAINEGONKPEIEVTWAKQGRTERLCSONPPVLKHHORBITTTTQSDQEE 1680
DB 1621 NACSNHAIAINEGONKPEIEVTWAKQGRTERLCSONPPVLKHHORBITTTTQSDQEE 1680
QY 1681 IDYDDTISVEKKKDFDIYDEDENQSPRSFOKTRHYFIAAVERLDWYGMSSSPHVLNR 1740
DB 1681 IDYDDTISVEKKKDFDIYDEDENQSPRSFOKTRHYFIAAVERLDWYGMSSSPHVLNR 1740
QY 1741 AQSGSVPOFKVQSFPTDGSFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASR 1800
DB 1741 AQSGSVPOFKVQSFPTDGSFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASR 1800
QY 1801 PYSVSSLSIYEEBQORQABPKKPFVZNEKTYFWKQHHMAPTKDFCDCKAWAYPSDV 1860
DB 1801 PYSVSSLSIYEEBQORQABPKKPFVZNEKTYFWKQHHMAPTKDFCDCKAWAYPSDV 1860
QY 1861 DLEKDVHSLGILPLLVCHTNTLNPAHQRCVTVQBFALPFTTFDETKSWYFTENMERNCRA 1920
DB 1861 DLEKDVHSLGILPLLVCHTNTLNPAHQRCVTVQBFALPFTTFDETKSWYFTENMERNCRA 1920
QY 1921 PCNTQMEDPFTKFNYPHAINGYIMDTLPGLVMAQDQQRIRWYLLSMGSENENIHSIHFGH 1980
DB 1921 PCNTQMEDPFTKFNYPHAINGYIMDTLPGLVMAQDQQRIRWYLLSMGSENENIHSIHFGH 1980
QY 1981 VFTVRK3EYKMALNLYLPGVFETVEMLPKAGIWRVECLIGEHLGAGMSTLFLVYSNKC 2040
DB 1981 VFTVRK3EYKMALNLYLPGVFETVEMLPKAGIWRVECLIGEHLGAGMSTLFLVYSNKC 2040

```

QY 2041 QTPLOMASCHIRDRQITASGOYGOWAPKLABLHSGSINAWSTKEPFWIKVDLLAPMII 2100
DB 2041 QTPLOMASCHIRDRQITASGOYGOWAPKLABLHSGSINAWSTKEPFWIKVDLLAPMII 2100
QY 2101 HGKIQGARQKFSSLIYISQFIIMYSLDGKKWQTYRGNSTGILMVFFGNDVSSGIKHNIEN 2160
DB 2101 HGKIQGARQKFSSLIYISQFIIMYSLDGKKWQTYRGNSTGILMVFFGNDVSSGIKHNIEN 2160
QY 2161 PPIIARYIRLHPHTHYSIRSTLRMLMGCDLNSCMLPGLMESKAISDAQITASSYFTNNMFA 2220
DB 2161 PPIIARYIRLHPHTHYSIRSTLRMLMGCDLNSCMLPGLMESKAISDAQITASSYFTNNMFA 2220
QY 2280 TWSPEKARHLQGRNANRPOVNNPEKMLQVDFQKTMKVGTGTVQGVKSLTSMYVKEFL 2280
DB 2280 TWSPEKARHLQGRNANRPOVNNPEKMLQVDFQKTMKVGTGTVQGVKSLTSMYVKEFL 2280
QY 2281 TSSSDGCHQWTLFFQNGKVKVQFGNQDGFTPVNSLDPLLTTRYLRIRHPQSWHQAIALRM 2340
DB 2281 TSSSDGCHQWTLFFQNGKVKVQFGNQDGFTPVNSLDPLLTTRYLRIRHPQSWHQAIALRM 2340
QY 2341 EVLGEAQDLY 2351
DB 2341 EVLGEAQDLY 2351

RESULT 3

US-08-366-851A-2
; Sequence 2, Application US/08366851A
; Patent No. 5681746
; GENERAL INFORMATION:
; APPLICANT: Bodner, Mordechai
; APPLICANT: De Polo, Nicolas J.
; APPLICANT: Hsu, David Chi-fang
; APPLICANT: Chang, Steven
; TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Viagene, Inc.
; STREET: 11055 Roselle Street
; CITY: San Diego
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,851A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Chambers, Daniel M.
; REGISTRATION NUMBER: 34,561
; REFERENCE/DOCKET NUMBER: 930049.438
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 452-1288
; TELEFAX: (619) 452-2616
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-366-851A-2
Query Match 99.8%; Score 12390; DB 1; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSDWYMQSDLGELPVDARFPFRPKSFPFN 60

DB 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSDWYMQSDLGELPVDARFPFRPKSFPFN 60
QY 61 TSVVYKKTFLVFEPTDHLFNIAKPRPPWMLGPTIQAEVYDVTVVITLKNMASHVSVSHAV 120
DB 61 TSVVYKKTFLVFEPTDHLFNIAKPRPPWMLGPTIQAEVYDVTVVITLKNMASHVSVSHAV 120
QY 121 GVSVMKASGAEYDDQTSOREKDDKVPFGSSHYVMQVLKENGPMASDPLCLTYSLSH 180
DB 121 GVSVMKASGAEYDDQTSOREKDDKVPFGSSHYVMQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFLLFAVPFDECKSWHSEFTKNSLMQORD 240
DB 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKFLLFAVPFDECKSWHSEFTKNSLMQORD 240
QY 241 AASARAPKMHVTNGYVNRSLPGLIGHKRSVYWHVIGMGTTPVHSIFLEGHTFLVRNH 300
DB 241 AASARAPKMHVTNGYVNRSLPGLIGHKRSVYWHVIGMGTTPVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITFLTAQTLMDLGOFLLSCHTSSSHOHGMEAYVKVDSCPEPPQIMKUNE 360
DB 301 ROASLEISPIITFLTAQTLMDLGOFLLSCHTSSSHOHGMEAYVKVDSCPEPPQIMKUNE 360
QY 361 EADYDDDLTDSMDVVRPDDNSPSPFIQIRSAVAKHPKTVWHYIAABEEDWDYAPLVLA 420
DB 361 EADYDDDLTDSMDVVRPDDNSPSPFIQIRSAVAKHPKTVWHYIAABEEDWDYAPLVLA 420
QY 421 PDRSYKSOYLNGPQIRGRYKVRFNAYTDEFKTEEAACHESGILGLLYGEVGTLL 480
DB 421 PDRSYKSOYLNGPQIRGRYKVRFNAYTDEFKTEEAACHESGILGLLYGEVGTLL 480
QY 481 LIIFKNOASRPVNIYPHCITDVRPLYSRRLPKGVHLDFFILPGEIFKYKWTVTVEDCP 540
DB 481 LIIFKNOASRPVNIYPHCITDVRPLYSRRLPKGVHLDFFILPGEIFKYKWTVTVEDCP 540
QY 541 TKSDPRCLTRYSSVNNMERDLASGLIGPLICYKESVDQKGNQIMSKRNVLFSVDE 600
DB 541 TKSDPRCLTRYSSVNNMERDLASGLIGPLICYKESVDQKGNQIMSKRNVLFSVDE 600
QY 601 NRSWYLTENIQRFENPAGVQLEDEPFOASNMHSGINGVVPDSLOSLVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQRFENPAGVQLEDEPFOASNMHSGINGVVPDSLOSLVCLHEVAYWYILS 660
QY 661 IGAQTDFLSVFSGYTFXHKWYVEDTLTFPFSGETVFMSEMENPGLWILGCHNSDFRNG 720
DB 661 IGAQTDFLSVFSGYTFXHKWYVEDTLTFPFSGETVFMSEMENPGLWILGCHNSDFRNG 720
QY 721 MTALLKVSSCDKNTGDYVEDSDYEDISAYLLSKNNAIEPRSPQNSRHRSTROKQFNATTI 780
DB 721 MTALLKVSSCDKNTGDYVEDSDYEDISAYLLSKNNAIEPRSPQNSRHRSTROKQFNATTI 780
QY 781 PENDIEKTDPPFAHRTPMPKIQNVASSDILLMLROSPTPHGLSLDLOBAKYETESDDPS 840
DB 781 PENDIEKTDPPFAHRTPMPKIQNVASSDILLMLROSPTPHGLSLDLOBAKYETESDDPS 840
QY 841 PGALDSNNLSSEWTHFRPOLHSGDMVTFPESGLQRLNEKLGTTAATLKKLDFKVSST 900
DB 841 PGALDSNNLSSEWTHFRPOLHSGDMVTFPESGLQRLNEKLGTTAATLKKLDFKVSST 900
QY 901 SNNLSTIPSDNLAAGTDNTSSLGPPSMVHVDSOLDTTLFGKKSPLTESGGPLSLSEE 960
DB 901 SNNLSTIPSDNLAAGTDNTSSLGPPSMVHVDSOLDTTLFGKKSPLTESGGPLSLSEE 960
QY 961 NNDSKLLSGLMNSQESSWGKNVSTESGRIFKGRAGFPALLTKDNALFKVSI SILLKTN 1020
DB 961 NNDSKLLSGLMNSQESSWGKNVSTESGRIFKGRAGFPALLTKDNALFKVSI SILLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSLLIENSFSVWQNIILESJTEFKKVTPLIHRMLMDKNATALEL 1080
DB 1021 KTSNNSATNRKTHIDGPSLLIENSFSVWQNIILESJTEFKKVTPLIHRMLMDKNATALEL 1080
QY 1081 NEWSNKTSSKNMVMVOQKEGFIIPDAQNPDMSPFKMLFLPESARWIOETHCKSLNSG 1140

1081 NHMSKNTSSKNMWMVQQKKEGPDPDAQNDMSFPMWLFLPESARVQRTGEGKNSLNSG 1140
1141 QGSPKQLVSLGPKSVGQNFISEKKNVVGKGEFTKQVGLKEMVFPSSRNPLPLTNLDN 1200
1141 QGSPKQLVSLGPKSVGQNFISEKKNVVGKGEFTKQVGLKEMVFPSSRNPLPLTNLDN 1200
1201 LHENNTNQKTKOEBTEKKEKTELLIQENNVLPQIHVTGTGNFMKNLFLLSRQNVESYD 1260
1201 LHENNTNQKTKOEBTEKKEKTELLIQENNVLPQIHVTGTGNFMKNLFLLSRQNVESYD 1260
1261 GAYAPVLQDFRSLNDSTNRKHTAHFSKKEEENLEGLNQTKQIVKCYACTTRISPT 1320
1261 GAYAPVLQDFRSLNDSTNRKHTAHFSKKEEENLEGLNQTKQIVKCYACTTRISPT 1320
1321 SQQNFVORSKALQKRLPLEETELKRIIIVDSTSTQMSKNMGLPSTLTQIDYNEKE 1380
1321 SQQNFVORSKALQKRLPLEETELKRIIIVDSTSTQMSKNMGLPSTLTQIDYNEKE 1380
1381 KGAITQSPLSCLTRSHSIPOANRSPPIAKVSSFPPIRPIYLRVLFDQNSSHLPAASY 1440
1381 KGAITQSPLSCLTRSHSIPOANRSPPIAKVSSFPPIRPIYLRVLFDQNSSHLPAASY 1440
1441 RKDQGVQESHFTQGAAXKNLSAILTLEMTGQREVSGISGTSATNSVTVKKVNTVLP 1500
1441 RKDQGVQESHFTQGAAXKNLSAILTLEMTGQREVSGISGTSATNSVTVKKVNTVLP 1500
1501 KPD:PKTSGKVLLPKVHIYQKDLFPPTETSNQSPCHLDLVEGSLQGTGEGAIKNEANRP 1560
1501 KPD:PKTSGKVLLPKVHIYQKDLFPPTETSNQSPCHLDLVEGSLQGTGEGAIKNEANRP 1560
1561 GKVPFLVATESSAKTSPKLDPLAWNDHYGTQIPKBEWKQESPEKTAFFKKDITLSL 1620
1561 GKVPFLVATESSAKTSPKLDPLAWNDHYGTQIPKBEWKQESPEKTAFFKKDITLSL 1620
1621 NACSNHAIAINQGNKPEIEVTWAKQGRTERLCSONPPVLKQHQRIETITLQSQOE 1680
1621 NACSNHAIAINQGNKPEIEVTWAKQGRTERLCSONPPVLKQHQRIETITLQSQOE 1680
1681 IDYDDTTSVENKKEDFDYDDEDNQSPRSFOKTRHYFIARAVERLDYGMSSSHVLNRP 1740
1681 IDYDDTTSVENKKEDFDYDDEDNQSPRSFOKTRHYFIARAVERLDYGMSSSHVLNRP 1740
1741 AQSGSVQPKKVVQFEPTDGSFTQPLRYGELNEHLGLLGPVIRAEVDNIMVTFRNQASR 1800
1741 AQSGSVQPKKVVQFEPTDGSFTQPLRYGELNEHLGLLGPVIRAEVDNIMVTFRNQASR 1800
1801 PYSYSSLSIYEBEORQGAEPKPKVKNETKTFWKVQHMAPTKDBFDCKAWAYPSDV 1860
1801 PYSYSSLSIYEBEORQGAEPKPKVKNETKTFWKVQHMAPTKDBFDCKAWAYPSDV 1860
1861 DLEKDVHSLGPIILLVCHTNTLNPAHQVQVTVQEPFALFPTTFDETKSWYFTENMERNCA 1920
1861 DLEKDVHSLGPIILLVCHTNTLNPAHQVQVTVQEPFALFPTTFDETKSWYFTENMERNCA 1920
1921 PCNTQMEDPTTKENYRFAHNGYIMDTLPGLVMAQDQIRWYLLSMGSGNENIHSHFSGH 1980
1921 PCNTQMEDPTTKENYRFAHNGYIMDTLPGLVMAQDQIRWYLLSMGSGNENIHSHFSGH 1980
1981 VFTVRKKEEYKVALYNLYPGVPEIVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKVALYNLYPGVPEIVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
2041 QTPLGMAHGSHRDFOITASGOVGOAPKLARLHYSGSNANSTKBPFPFWIYVDLLAPMII 2100
2041 QTPLGMAHGSHRDFOITASGOVGOAPKLARLHYSGSNANSTKBPFPFWIYVDLLAPMII 2100
2101 HGKIQGARQKPFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMWFFGNVDSGKHNIFN 2160
2101 HGKIQGARQKPFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMWFFGNVDSGKHNIFN 2160
2161 PPIIARYIRLHPHTHYSIRSLRMELMGDLNSCSMPLGMESKAIISDAQITASSYFTNMFA 2220
2161 PPIIARYIRLHPHTHYSIRSLRMELMGDLNSCSMPLGMESKAIISDAQITASSYFTNMFA 2220

QY 2221 TWSPSKARLHLQGRSNARPOVANNPKEWLQVDFQTKMKVTGVTTOGVKSLTSMYVKBEFL 2380
Db 2221 TWSPSKARLHLQGRSNARPOVANNPKEWLQVDFQTKMKVTGVTTOGVKSLTSMYVKBEFL 2380
QY 2281 ISSSQDGHQWTLFFQNGKVKVFGQNGDSFTFVNVNSLDPPILLTRYLRHPHOSWVHQIALRM 2340
Db 2281 ISSSQDGHQWTLFFQNGKVKVFGQNGDSFTFVNVNSLDPPILLTRYLRHPHOSWVHQIALRM 2340
QY 2341 EVLGCEAQQDLY 2351
Db 2341 EVLGCEAQQDLY 2351

RESULT 4

US-08-121-202-2
; Sequence 2, Application US/09121202
; Patent No. 5563045
; GENERAL INFORMATION:
; APPLICANT: Pittman, Debra
; APPLICANT: Rehemtulla, Alinawaz
; APPLICANT: Wozney, John M.
; APPLICANT: Kaufman, Randal J.
; TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: JS/08/121,202
; FILING DATE: 14-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, M. C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1210 X8574
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2351 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-121-202-2

Query Match 99.8%; Score 12386; DB 1; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2346; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MQIELSTCFCLLRFCSATRRYYLGAVELSWDMQSDLGELPVDARFPPRPVPSFPFN 60
Db 1 MQIELSTCFCLLRFCSATRRYYLGAVELSWDMQSDLGELPVDARFPPRPVPSFPFN 60
QY 61 TSVYVKKTLFVEFTDHLFNIAKPEPPMGLGPTIOAEVYDVTVTILKNKASHPVSLHAV 120
Db 61 TSVYVKKTLFVEFTDHLFNIAKPEPPMGLGPTIOAEVYDVTVTILKNKASHPVSLHAV 120
QY 121 GVSWKASGEAYDDQTSQREKEDDKVFPGSGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSWKASGEAYDDQTSQREKEDDKVFPGSGSHYVWQVLKENGPMASDPLCLTYSYLSH 180
QY 181 VDLVKDNLNSGLIGALLVCREGSLAKEKTQTLHKPILLFAVDEGKSWHSETYNSLMQDRD 240

Db 181 VDLVKOLNSGLIGALLVCREGSLAKEKQTQLHKPILFAVFEDEKSNHSETKNSLMQDRD 240
Qy 241 AASARAWPMTHTVNGYVNRSLPGLIGCHRKVYVHVHVGCMGTTPVHVSIFLEGHTFLVRNH 300
Db 241 AASARAWPMTHTVNGYVNRSLPGLIGCHRKVYVHVHVGCMGTTPVHVSIFLEGHTFLVRNH 300
Qy 301 ROASLEISPIFLTAQTLMLDGLFLLSCHTSSHOHDMQEAIVKVDSCPEBPQIMKONE 360
Db 301 ROASLEISPIFLTAQTLMLDGLFLLSCHTSSHOHDMQEAIVKVDSCPEBPQIMKONE 360
Qy 361 EAEYDDDLTDSMDVVRFDONSPESFQIORSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
Db 361 EAEYDDDLTDSMDVVRFDONSPESFQIORSVAKKHPTWVHYIAABEEDWDYAPLVLA 420
Qy 421 PDRSRYSKOYLLNGHQRTGRKVKYKTRFMAYTDETEKTEBAAOHBESGILGLLYGEVGTIL 480
Db 421 PDRSRYSKOYLLNGHQRTGRKVKYKTRFMAYTDETEKTEBAAOHBESGILGLLYGEVGTIL 480
Qy 481 LIIIFKQASRPNIYVPHGTDVRLYSRRLPKGVKHLXDFPILPGEIEFKYKWTVTVEDCP 540
Db 481 LIIIFKQASRPNIYVPHGTDVRLYSRRLPKGVKHLXDFPILPGEIEFKYKWTVTVEDCP 540
Qy 541 TKSDPRCLTRYSSFWNMRDLASGLIGELLYCYKESVDQKQNMSCRNVILFSVEDE 600
Db 541 TKSDPRCLTRYSSFWNMRDLASGLIGELLYCYKESVDQKQNMSCRNVILFSVEDE 600
Qy 601 NRSWYLTENIQRFLENPAGVQLEDPEFOASNIHMSINGVVDPSLOLSVCLHEVAVWYILS 660
Db 601 NRSWYLTENIQRFLENPAGVQLEDPEFOASNIHMSINGVVDPSLOLSVCLHEVAVWYILS 660
Qy 661 IGAQTDFLSVFSGYTFXHKHMYEYDZTLILFPFSGTVMFSMENPGLWILGCHNSDFNRNG 720
Db 661 IGAQTDFLSVFSGYTFXHKHMYEYDZTLILFPFSGTVMFSMENPGLWILGCHNSDFNRNG 720
Qy 721 MTALLKVSQCDKNTGYEDSDYEDISAYLLSKNNAIEPRSPQNSKHPSSTROKQNAATTI 780
Db 721 MTALLKVSQCDKNTGYEDSDYEDISAYLLSKNNAIEPRSPQNSKHPSSTROKQNAATTI 780
Qy 781 PENDIEKTDPEFAHRTBPKONVASSDLMLLROSPTPHGLSLSDLOEAKYETESDDPS 840
Db 781 PENDIEKTDPEFAHRTBPKONVASSDLMLLROSPTPHGLSLSDLOEAKYETESDDPS 840
Qy 841 PGAISSNLSLSEWTHFRPOLHESGDMVFTPEGLQLRLNEKLGTAAATELKKLDFKVSST 900
Db 841 PGAISSNLSLSEWTHFRPOLHESGDMVFTPEGLQLRLNEKLGTAAATELKKLDFKVSST 900
Qy 901 SNNLITIPSDMLAAGTDNTSSLGPPSPVHYDSQDITTLFQKKSPLTESGGPLSLSEE 960
Db 901 SNNLITIPSDMLAAGTDNTSSLGPPSPVHYDSQDITTLFQKKSPLTESGGPLSLSEE 960
Qy 961 NNDKSLLESGLMNSQESSWGKNVSTTESGRLFKGRAGHPALLTXDNALFKVSIISLLKTN 1020
Db 961 NNDKSLLESGLMNSQESSWGKNVSTTESGRLFKGRAGHPALLTXDNALFKVSIISLLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESOTEFKKVTPLIHDRMLMDKNATAILRL 1080
Db 1021 KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESOTEFKKVTPLIHDRMLMDKNATAILRL 1080
Qy 1081 NEMSNTTSSKNEMVQOKKEGPIPPDAONPMSPEKMLFLPESAEWIORTHCKNSLNSG 1140
Db 1081 NEMSNTTSSKNEMVQOKKEGPIPPDAONPMSPEKMLFLPESAEWIORTHCKNSLNSG 1140
Qy 1141 QGPPSKQLVSLGPEKVEGONFLSEKKNVVGKGETTKOVLKEMVFPSSRNLFITNLND 1200
Db 1141 QGPPSKQLVSLGPEKVEGONFLSEKKNVVGKGETTKOVLKEMVFPSSRNLFITNLND 1200
Qy 1201 LHENNTHNQEKTIQEBIEKKEKTLIQENVVLPQHTVTGTGNFMKXNLLFLSTRQNVGSYD 1260
Db 1201 LHENNTHNQEKTIQEBIEKKEKTLIQENVVLPQHTVTGTGNFMKXNLLFLSTRQNVGSYD 1260
Qy 1261 GAYAVLQDFRSLNDSTNRKTKHTAHSKKGSEENLEGNGNOTKQIVKCYACTRISPNPT 1320

Db 1321 GAVAVLQDFRSLNDSTNRKTKHTAHSKKGSEENLEGNGNOTKQIVKCYACTRISPNPT 1320
Qy 1321 SQQNFVTOQRKRALQKQRLP-EETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFVTOQRKRALQKQRLP-EETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGAIQTQSPLSDCTRSHSIPOANRSPPIAKVSSFPPIRPIYITRVLPODNSSHLPAASY 1440
Db 1381 KGAIQTQSPLSDCTRSHSIPOANRSPPIAKVSSFPPIRPIYITRVLPODNSSHLPAASY 1440
Qy 1441 RKDQSGVQESHSFLOGAKKNLSAILITLMTQDQREVSGLSATNSVTKYKVENTVLP 1500
Db 1441 RKDQSGVQESHSFLOGAKKNLSAILITLMTQDQREVSGLSATNSVTKYKVENTVLP 1500
Qy 1501 KPDLPKTSKGVKVELLPKHVHYKOLPPTETSNKSGPHGLDVEGSLLOQTEGAIKWNEANRP 1560
Db 1501 KPDLPKTSKGVKVELLPKHVHYKOLPPTETSNKSGPHGLDVEGSLLOQTEGAIKWNEANRP 1560
Qy 1561 GKVPFLRVATESSAKTSPKLLDPLAWNHYGTQPKBEWKSQESPEKTAPEKKDITLSL 1620
Db 1561 GKVPFLRVATESSAKTSPKLLDPLAWNHYGTQPKBEWKSQESPEKTAPEKKDITLSL 1620
Qy 1621 NACSNHAIAINBGONKPELEVTWAKQGRTERLCSQNPVULKBOREILTATTI-QSQOEE 1680
Db 1621 NACSNHAIAINBGONKPELEVTWAKQGRTERLCSQNPVULKBOREILTATTI-QSQOEE 1680
Qy 1681 IDYDDTISVENMKKEDFDIYDEDNQSPESPOKTRHYFIAAVERLWDYMGSSSPEVLRNR 1740
Db 1681 IDYDDTISVENMKKEDFDIYDEDNQSPESPOKTRHYFIAAVERLWDYMGSSSPEVLRNR 1740
Qy 1741 AQSQSVPOFKKVQFQBTDDGFTQPLYGELNEHLGLLGPYIRAEVDNIMWTFRNOASR 1800
Db 1741 AQSQSVPOFKKVQFQBTDDGFTQPLYGELNEHLGLLGPYIRAEVDNIMWTFRNOASR 1800
Qy 1801 PYSYSSLIISYEEORQGAEPKQNFVKNETKTYFWKVQOHHMPTKDBFCCKAWAYPSDV 1860
Db 1801 PYSYSSLIISYEEORQGAEPKQNFVKNETKTYFWKVQOHHMPTKDBFCCKAWAYPSDV 1860
Qy 1861 DLEKDVHSGLLGPLLVCHTNTLNPAGHQVTVQBFALFTTIFDETKSWYFTENMERNCA 1920
Db 1861 DLEKDVHSGLLGPLLVCHTNTLNPAGHQVTVQBFALFTTIFDETKSWYFTENMERNCA 1920
Qy 1921 PCNIQMEDPTPKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHFSGH 1980
Db 1921 PCNIQMEDPTPKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSGNENIHSIHFSGH 1980
Qy 1981 VFTVRKKEEYKXALYNLYPGYFETVEMLPKAGIWRVECLICEHLHAGWSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKXALYNLYPGYFETVEMLPKAGIWRVECLICEHLHAGWSTLFLVYSNKC 2040
Qy 2041 QTPGLMASGHIRDFOITASGOYGQWAPKLARLHYSGSINAMSTKEPFWIKVDLLAPMII 2100
Db 2041 QTPGLMASGHIRDFOITASGOYGQWAPKLARLHYSGSINAMSTKEPFWIKVDLLAPMII 2100
Qy 2101 HGIKTCGAROKFESSLYISQFIIMYSLDGKKWQTYRGNSTGTLVWFFGNVDSSGKHNLN 2160
Db 2101 HGIKTCGAROKFESSLYISQFIIMYSLDGKKWQTYRGNSTGTLVWFFGNVDSSGKHNLN 2160
Qy 2161 PPIIARYIRLHPTVYSIRSTLRMELMGCDLNSCSMPLGWESKAI-SDAQITASSYFTNFA 2220
Db 2161 PPIIARYIRLHPTVYSIRSTLRMELMGCDLNSCSMPLGWESKAI-SDAQITASSYFTNFA 2220
Qy 2221 TWSPSKAEHLHQGSNAWRPQVNNPKWLQVDFQKTMKVTVGTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKAEHLHQGSNAWRPQVNNPKWLQVDFQKTMKVTVGTQGVKSLTSMYVKEFL 2280
Qy 2281 TSSSQDGHQWTLFPQNGKVKVPOGQNDSTFTVWNSLDDPILTRYLRIHPOSWHQAIAEM 2340
Db 2281 TSSSQDGHQWTLFPQNGKVKVPOGQNDSTFTVWNSLDDPILTRYLRIHPOSWHQAIAEM 2340
Qy 2341 EVLIGCEAODLY 2351
Db 2341 EVLIGCEAODLY 2351


```
RESULT 5
5422260-1
;PATENT NO. 5422260
;APPLICANT: KAUFMAN, RANDAL J.;PITTMAN, DEBRA D.;TOOLE, JOHN J.
;TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTEINS
;NUMBER OF SEQUENCES: 15
;CURRENT APPLICATION DATA:
;APPLICATION NUMBER: US/07/883,936
;FILING DATE: 15-MAY-1992
;PRIOR APPLICATION DATA:
;APPLICATION NUMBER: 279,485
;FILING DATE: 02-DEC-1988; 09-DEC-1986
;APPLICATION NUMBER: 939,658
;FILING DATE: 09-DEC-1986
;APPLICATION NUMBER: 932,767
;FILING DATE: 18-NOV-1986
;APPLICATION NUMBER: 868,410
;FILING DATE: 29-MAY-1986
;SEQ ID NO:1:
;LENGTH: 2351
5422260-1

Query Match          99.7%; Score 12377; DB 6; Length 2351;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2345; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MQIELSTCFPLCLLRFCFSATRRYVLGAVELSDWMQSDLGELPVDARFPFRVPSKSPFN 60
Db 1 MQIELSTCFPLCLLRFCFSATRRYVLGAVELSDWMQSDLGELPVDARFPFRVPSKSPFN 60

Qy 61 TSVVYKKTLEVFETDHLFNIAKPPNMGILGPLOAIVYTVITLKNMASHPVSLHAV 120
Db 61 TSVVYKKTLEVFETDHLFNIAKPPNMGILGPLOAIVYTVITLKNMASHPVSLHAV 120

Qy 121 GVSYWKASEGAYDDOTSQRKEDDKVPPGSGHTYVWQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSYWKASEGAYDDOTSQRKEDDKVPPGSGHTYVWQVLKENGPMASDPLCLTYSYLSH 180

Qy 181 VDLVKDLSGLIGALLVCRBGLAKETQTLHKFILLFAVFDGKSHWSTKNSLMQDRD 240
Db 181 VDLVKDLSGLIGALLVCRBGLAKETQTLHKFILLFAVFDGKSHWSTKNSLMQDRD 240

Qy 241 AASARAWPMHTVNGYVNRSLPGLTGCHRSKYVHWIGMTPEVHSIFLEGHTFLVRNH 300
Db 241 AASARAWPMHTVNGYVNRSLPGLTGCHRSKYVHWIGMTPEVHSIFLEGHTFLVRNH 300

Qy 301 RQASLEISPIITPLTAQTLIMDLGQFLLSCHISSHQHGDGMEAYVKVDSCPEBPQLMKNE 360
Db 301 RQASLEISPIITPLTAQTLIMDLGQFLLSCHISSHQHGDGMEAYVKVDSCPEBPQLMKNE 360

Qy 361 EAEDYDDDLTDSMDVVRPDDNNSPFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420
Db 361 EAEDYDDDLTDSMDVVRPDDNNSPFTQIRSVAKKPKTWVHYIAAEEEDWDYAPLVLA 420

Qy 421 PDRSYKQYLNNGPQIRGRKVKYVRMAYTDETFKTRAIQHBGSLGLPGLLYGVGDTL 480
Db 421 PDRSYKQYLNNGPQIRGRKVKYVRMAYTDETFKTRAIQHBGSLGLPGLLYGVGDTL 480

Qy 481 LLIIFKQASRPNIYPHGIDTVRPPYSRRLPKGVKHLKDFPILPGEIPIKYKWTVVEDGP 540
Db 481 LLIIFKQASRPNIYPHGIDTVRPPYSRRLPKGVKHLKDFPILPGEIPIKYKWTVVEDGP 540

Qy 541 TKSDPRCLTRYSSFNWBERDLASGLIGPLLICYKESVDQKGNQIMSKRNWILSVFDE 600
Db 541 TKSDPRCLTRYSSFNWBERDLASGLIGPLLICYKESVDQKGNQIMSKRNWILSVFDE 600

Qy 601 NESWYLTENIQRLPNPAGVQLDEPFOASNMWESINGVVPDSLQSVCLHVAWYIILS 660
Db 601 NESWYLTENIQRLPNPAGVQLDEPFOASNMWESINGVVPDSLQSVCLHVAWYIILS 660

Qy 661 IGAQTDFLSVFSGYTFKHVMYEDTLTLPFSGETVFMENPGLWILGCHNSDFNRG 720
Db 661 IGAQTDFLSVFSGYTFKHVMYEDTLTLPFSGETVFMENPGLWILGCHNSDFNRG 720
```

```
Db 661 IGAQTDFLSVFSGYTFKHVMYEDTLTLPFSGETVFMENPGLWILGCHNSDFNRG 720
Qy 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTPQKQFNATTI 780
Db 721 MTALLKVSSCDKNTGDIYEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTPQKQFNATTI 780
Qy 781 PENDIEKTDPEAHRTPMPKIQNVSSDDLMLLRQSPTPHGLSLSDLOQAEKYEFSDPS 840
Db 781 PENDIEKTDPEAHRTPMPKIQNVSSDDLMLLRQSPTPHGLSLSDLOQAEKYEFSDPS 840
Qy 841 PGAIDSNNSLSEMTFRPQLHHSQDMVFTPESGIQLRNEKLGTTAATELKKLDFKVSST 900
Db 841 PGAIDSNNSLSEMTFRPQLHHSQDMVFTPESGIQLRNEKLGTTAATELKKLDFKVSST 900
Qy 901 SNNLISITIPSDNLAAGTDNTSSILGPPMPVHYDSOLDTTLFGKSSPLTSGGFLSSEE 960
Db 901 SNNLISITIPSDNLAAGTDNTSSILGPPMPVHYDSOLDTTLFGKSSPLTSGGFLSSEE 960
Qy 961 NNDKLLIESGLMNSQESSWGKNVSSSTESGRLEFKGRAHGPAALLTKDNALFKVSTSLKTN 1020
Db 961 NNDKLLIESGLMNSQESSWGKNVSSSTESGRLEFKGRAHGPAALLTKDNALFKVSTSLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGFSLLIENSPPVWQNLLESDETFKKVTPLIHDMMLMDKNATAURL 1080
Db 1021 KTSNNSATNRKTHIDGFSLLIENSPPVWQNLLESDETFKKVTPLIHDMMLMDKNATAURL 1080
Qy 1081 NHMKNKTTSSKNMHWQCKEGPIPPDAQNPDSFFKMLZLPESARWLTQRTGKNSLNSG 1140
Db 1081 NHMKNKTTSSKNMHWQCKEGPIPPDAQNPDSFFKMLZLPESARWLTQRTGKNSLNSG 1140
Qy 1141 QGPPKQNLISLGPESKVEGQNFVSEKNKVVVGKEFTKDVGLKEMVFPSSRNLFITLND 1200
Db 1141 QGPPKQNLISLGPESKVEGQNFVSEKNKVVVGKEFTKDVGLKEMVFPSSRNLFITLND 1200
Qy 1201 LHENNTNNOEKKIQEEIEKKTETLLOENWVLPQIHVTGTGKNFMKNLFPILLSTRQNVESYE 1260
Db 1201 LHENNTNNOEKKIQEEIEKKTETLLOENWVLPQIHVTGTGKNFMKNLFPILLSTRQNVESYE 1260
Qy 1261 GAYAPVLQDPRSLNDSINRIKHTAHFSKKEENLEGLGNQTKQIVEKYACTTRISNT 1320
Db 1261 GAYAPVLQDPRSLNDSINRIKHTAHFSKKEENLEGLGNQTKQIVEKYACTTRISNT 1320
Qy 1321 SQQNFVQSRKALKQRLPLEETELEKRIITVDSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFVQSRKALKQRLPLEETELEKRIITVDSTQWSKNMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGAITQSPSLDCLTRSHSIPOANRSPPIAKVSSPPIRPIYLTRVLPODNSSHLPAASY 1440
Db 1381 KGAITQSPSLDCLTRSHSIPOANRSPPIAKVSSPPIRPIYLTRVLPODNSSHLPAASY 1440
Qy 1441 RKDQSGVQESSHFLQGAKKNNLSLAULTLEMTGQREVSGISGTSATNSVTYKVENTVLP 1500
Db 1441 RKDQSGVQESSHFLQGAKKNNLSLAULTLEMTGQREVSGISGTSATNSVTYKVENTVLP 1500
Qy 1501 KPDLPKTSKGVKELLPKVHIYQKDLFPPTETSNSSPGHLDLVEGSLQGTGEGAIKWEANRP 1560
Db 1501 KPDLPKTSKGVKELLPKVHIYQKDLFPPTETSNSSPGHLDLVEGSLQGTGEGAIKWEANRP 1560
Qy 1561 GKVPFLRVATSSAKTPSKLLDPLAWNHGYTQIPKEWKQKESPEKTAFFKKDTLISL 1620
Db 1561 GKVPFLRVATSSAKTPSKLLDPLAWNHGYTQIPKEWKQKESPEKTAFFKKDTLISL 1620
Qy 1621 NACSNHAIAINSGONKPEIEVTWAKGRTERLCSQNPVVKRHHOREITRTTILQSDOE 1680
Db 1621 NACSNHAIAINSGONKPEIEVTWAKGRTERLCSQNPVVKRHHOREITRTTILQSDOE 1680
Qy 1681 IDYDDTIQSVEMKKEDFDIYDSDENQSPRSFQKTRHYPIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTIQSVEMKKEDFDIYDSDENQSPRSFQKTRHYPIAAVERLWDYGMSSSPHVLNR 1740
Qy 1741 AQSQSVPOFKKVCQFTDQSGFTQPLYEGELNEHLGLLGPYIRAEVDNIMVTFRQASR 1800
Db 1741 AQSQSVPOFKKVCQFTDQSGFTQPLYEGELNEHLGLLGPYIRAEVDNIMVTFRQASR 1800
```

1741 AQSQSVQPKVKKVQFETPDGSGFTQPLRYGELNEHLGLGPVIRAEVEDNIMVTFRNQASR 1800
1801 PYSYSSLSIYEEBQROGAERKPKVFNENKTYFWKQCHMAPTKDEFDCANAYPSDV 1860
1801 PYSYSSLSIYEEBQROGAERKPKVFNENKTYFWKQCHMAPTKDEFDCANAYPSDV 1860
1861 DLEKDVHSLGLPLLVCHTNTLNAFHGRQVTVQFALPFTTFDETKSWYFTENNERCRA 1920
1861 DLEKDVHSLGLPLLVCHTNTLNAFHGRQVTVQFALPFTTFDETKSWYFTENNERCRA 1920
1921 PCNTQMEDPTKPNYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIFSGH 1980
1921 PCNTQMEDPTKPNYRFAHNGYIMDTLPLGLVMAQDQIRWYLLSMGNSNENIHSIFSGH 1980
1981 VFTVRKKEEYKVALYNLPYGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
1981 VFTVRKKEEYKVALYNLPYGVFETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
2041 QTPGLMASGHIRDFOITASGOYGQWAPKRLARLHYSGSINAWSTKEPFSWIKVDLIAPMII 2100
2041 QTPGLMASGHIRDFOITASGOYGQWAPKRLARLHYSGSINAWSTKEPFSWIKVDLIAPMII 2100
2101 HGKITQGARQKPSLSYISQFIIMYSLDGKKQWQTYRGNSTGILMVPFGNVDSGKHNIFN 2160
2101 HGKITQGARQKPSLSYISQFIIMYSLDGKKQWQTYRGNSTGILMVPFGNVDSGKHNIFN 2160
2161 PPIIARYRLRPHYTHSIRSTRLMELMGDLNCSMPLGNESKAISDAQITASSYFTNMFA 2220
2161 PPIIARYRLRPHYTHSIRSTRLMELMGDLNCSMPLGNESKAISDAQITASSYFTNMFA 2220
2221 TWSPSKARLHLQGRSNARWPQNNPKWLQVDFQKTMKVGTGTVQVKSLLTSMYVKEFL 2280
2221 TWSPSKARLHLQGRSNARWPQNNPKWLQVDFQKTMKVGTGTVQVKSLLTSMYVKEFL 2280
2281 TSSSDQGHQWTLFFQNGKVKVQKQDSFTPVNSLDPLLRILRIHPQSWVHQIALRM 2340
2281 TSSSDQGHQWTLFFQNGKVKVQKQDSFTPVNSLDPLLRILRIHPQSWVHQIALRM 2340
2341 EVLGC EAQDLY 2351
2341 EVLGC EAQDLY 2351

RESULT 6

US-07-864-004B-4
; Sequence 4, Application US/07864004B
; Patent No. 5364771
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/864,004B
; FILING DATE: 07 APRIL 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-815-6508
TELEFAX: 404-815-6855
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
US-07-864-004B-4
Query Match 98.9%; Score 12282; DB 1; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRSYVYLGAVELSDYMSQDLGELPVDARPPRPVKSPPTNTSVVYKTLFVEFTDHLFN 79
DB 1 ATRYYVYLGAVELSDYMSQDLGELPVDARPPRPVKSPPTNTSVYKTLFVEFTDHLFN 60
QY 80 IAKPRPPMGLLPTIOAEVYDVTVITLKNVASHPVSLHAVGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPMGLLPTIOAEVYDVTVITLKNVASHPVSLHAVGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVFPGGSTTYVWVLKENGPMASDPLCTYSYLSVLDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVFPGGSTTYVWVLKENGPMASDPLCTYSYLSVLDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTOFLHPIILLFAVDFDEGKSWHSTKNSLMODRDAAGARAWPKMHTVNGYVNR 259
DB 191 EGS LAKEKTOFLHPIILLFAVDFDEGKSWHSTKNSLMODRDAAGARAWPKMHTVNGYVNR 240
QY 260 SLPLGLIGCHRSKVYVHHVIGMCTTPVHSHIFLEGHTFLVRNHRQASLEISP-TFLTAQTLL 319
DB 241 SLPLGLIGCHRSKVYVHHVIGMCTTPVHSHIFLEGHTFLVRNHRQASLEISP-TFLTAQTLL 300
QY 320 MDLQFLLSCHISSHQHDMGMEAYVKVDCSPEPOLIMKNNEEADYDDDLTDSMDVVRP 379
DB 301 MDLQFLLSCHISSHQHDMGMEAYVKVDCSPEPOLIMKNNEEADYDDDLTDSMDVVRP 360
QY 380 DDNSPSPFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVAPLAPDDRYSYKSOYLNNGPQIG 439
DB 361 DDNSPSPFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVAPLAPDDRYSYKSOYLNNGPQIG 420
QY 440 RYKVKVRPMAYTDETFKTRAIQHESGILGPLLVEVGDTLLIIPKQASRPYNYIPHGI 499
DB 421 RYKVKVRPMAYTDETFKTRAIQHESGILGPLLVEVGDTLLIIPKQASRPYNYIPHGI 480
QY 500 TDVREPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVTVEDGPKSDPRCLTYYSYFVNM 559
DB 481 TDVREPLYSRRLPKGVKHLKDFPILPGEIFPKYKWTVTVEDGPKSDPRCLTYYSYFVNM 540
QY 560 RDLASGLIGPLLIICYKESVDQKNGQIMGDKRNVILFSVFDENRSMYLTENIQRFLENPAG 619
DB 541 RDLASGLIGPLLIICYKESVDQKNGQIMGDKRNVILFSVFDENRSMYLTENIQRFLENPAG 600
QY 620 VOLEDPFQASNINHSINGYVFDLSIQLSVCLHEVAYWYILSIGACTDFLSVFFSGYTFKH 679
DB 601 VOLEDPFQASNINHSINGYVFDLSIQLSVCLHEVAYWYILSIGACTDFLSVFFSGYTFKH 660
QY 680 KMVYEDTLTLFPFSGETVMSXENPGLMLLGCHNSDFNRGMWTLKLYSSCDKNTGDYVE 739
DB 661 KMVYEDTLTLFPFSGETVMSXENPGLMLLGCHNSDFNRGMWTLKLYSSCDKNTGDYVE 720
QY 740 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTFQKQFNATTTIPENDIEKTDPAFAHRTMP 799
DB 721 DSYEDISAYLLSKNNAIEPRSPQNSRHPSTFQKQFNATTTIPENDIEKTDPAFAHRTMP 780

QY 800 KIQNVSSDILLMLLRQPTPHGLSLSDQAKYETTFSDPSPGALDSNNSLSEMTHERPQ 859
DB 781 KIQNVSSDILLMLLRQPTPHGLSLSDQAKYETTFSDPSPGALDSNNSLSEMTHERPQ 840
QY 860 LHHGDMVFTPESGLOJRLNKLGAATATELKKLDFKVSSTSNLLISTIPSDNLAAGTDN 919
DB 841 LHHGDMVFTPESGLOJRLNKLGAATATELKKLDFKVSSTSNLLISTIPSDNLAAGTDN 900
QY 920 TSSLGPPMPHYDSOLDTTLFGKXSPLETSBGGPLSLSEBNDKSKLLESGLMNSQESSW 979
DB 901 TSSLGPPMPHYDSOLDTTLFGKXSPLETSBGGPLSLSEBNDKSKLLESGLMNSQESSW 960
QY 980 GKNVSSPESGFLFKGKRAHGPAALTKONALFKVSISSLKTKNKTSNNSATNKTTHIDGPSL 1039
DB 961 GKNVSSPESGFLFKGKRAHGPAALTKONALFKVSISSLKTKNKTSNNSATNKTTHIDGPSL 1020
QY 1040 LIENSPPVQNLLESDETFEKKVTPLIHDMRLMDKNATRLRNHMSNKTTSKKNMEMVQOK 1099
DB 1021 LIENSPPVQNLLESDETFEKKVTPLIHDMRLMDKNATRLRNHMSNKTTSKKNMEMVQOK 1080
QY 1100 KEGPIPPDAQNPMDSFFKXFLPESARWIQPTHGKNSLNSGGSPKQOLVSLGPEKSVEG 1159
DB 1081 KEGPIPPDAQNPMDSFFKXFLPESARWIQPTHGKNSLNSGGSPKQOLVSLGPEKSVEG 1140
QY 1160 QNFLSEKNKVVGGEFTKDVGLKEMVPPSPRNFLPLTNLNLHNNTHNQSKKIQEBIEK 1219
DB 1141 QNFLSEKNKVVGGEFTKDVGLKEMVPPSPRNFLPLTNLNLHNNTHNQSKKIQEBIEK 1200
QY 1220 KETLIOENVVLPOHTVTGTQKNFMKNLELLSTRONVEGSDGAYAPVLQDPRSLNDSTNR 1279
DB 1201 KETLIOENVVLPOHTVTGTQKNFMKNLELLSTRONVEGSDGAYAPVLQDPRSLNDSTNR 1260
QY 1280 TKKHTAHFKSGEBENLEGLNQTKQIVEXYACVATRIISPNTSQONFVQSRKALKOPRL 1339
DB 1261 TKKHTAHFKSGEBENLEGLNQTKQIVEXYACVATRIISPNTSQONFVQSRKALKOPRL 1320
QY 1340 PLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITOSPLSDCLTRSHSI 1399
DB 1321 PLEETELEKRIIVDDTSTQWKNMKHLPSTLTQIDYNEKEKGAITOSPLSDCLTRSHSI 1380
QY 1400 POANRSLPIAKVSPSIRIPIYITRVLPONSHLPAASTRKDSGVQSRSHFLQAKK 1459
DB 1381 POANRSLPIAKVSPSIRIPIYITRVLPONSHLPAASTRKDSGVQSRSHFLQAKK 1440
QY 1460 NNLSLAILTLEMTGDQREVGSLGTSATNSVTKKVENTVLKPDLPKTSKGVLELLPKVHI 1519
DB 1441 NNLSLAILTLEMTGDQREVGSLGTSATNSVTKKVENTVLKPDLPKTSKGVLELLPKVHI 1500
QY 1520 YQKOLFPFTETNSGPHGLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1579
DB 1501 YQKOLFPFTETNSGPHGLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESAKTPSK 1560
QY 1580 LLDPLANDNHYGTOIPKEWKSQKSPKTAFFKDDTILSNACESHAIINAGQNK 1639
DB 1561 LLDPLANDNHYGTOIPKEWKSQKSPKTAFFKDDTILSNACESHAIINAGQNK 1620
QY 1640 BIEVTWAKQRTERLCSNPPVLKPHQSEIRITTLQSDQESIDYDDTISVEMKEDFDIY 1699
DB 1621 BIEVTWAKQRTERLCSNPPVLKPHQSEIRITTLQSDQESIDYDDTISVEMKEDFDIY 1680
QY 1700 DEDNQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNRNQAQSGVFPQKVVQFQFTD 1759
DB 1681 DEDNQSPRSFQKTRHYFIAAVERLMDYGMSSSPHVLNRNQAQSGVFPQKVVQFQFTD 1740
QY 1760 GSFTQPLVYRGELNHLGLGPIYIAEVEDNMTWFRNQAQSPYFYSLSLIEYEDQOQA 1819
DB 1741 GSFTQPLVYRGELNHLGLGPIYIAEVEDNMTWFRNQAQSPYFYSLSLIEYEDQOQA 1800
QY 1820 EPRKNFVNKTXYFWKVQHHMAPTKDEPCKAWAYPSDVLKDVHSGHLIGLILVCHT 1879
DB 1801 EPRKNFVNKTXYFWKVQHHMAPTKDEPCKAWAYPSDVLKDVHSGHLIGLILVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHA 1939

DB 1861 NTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHA 1920
QY 1940 INGVIMDTLPLGVMAQDQIRWYLLSMGNSNIHSHFSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGVIMDTLPLGVMAQDQIRWYLLSMGNSNIHSHFSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVETVEMLSKAGIWEVEGLIGBHLFAGMSTPLVYSNKCOTPLGMASGHIRDOITAS 2059
DB 1981 GVETVEMLSKAGIWEVEGLIGBHLFAGMSTPLVYSNKCOTPLGMASGHIRDOITAS 2040
QY 2060 GOYQOWAPKLARLHYSGSINAWSTKEPFSWKVLLAPMIHGIKTQCARQKPSLSYISQ 2119
DB 2041 GOYQOWAPKLARLHYSGSINAWSTKEPFSWKVLLAPMIHGIKTQCARQKPSLSYISQ 2100
QY 2120 FIIMYSLDGKKWQTYRGNSTGTMLVFFGNVDSSGIKHINFPPIIARIYIRLHPHYIRS 2179
DB 2101 FIIMYSLDGKKWQTYRGNSTGTMLVFFGNVDSSGIKHINFPPIIARIYIRLHPHYIRS 2160
QY 2180 TLRMELMGCDLNSGSMPLGMESKAIQITASSYFTNMFEATWSPSKARLHLQGRSNAR 2239
DB 2161 TLRMELMGCDLNSGSMPLGMESKAIQITASSYFTNMFEATWSPSKARLHLQGRSNAR 2220
QY 2240 PQVNNPKEMQLQVDPQKMKVKTGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 PQVNNPKEMQLQVDPQKMKVKTGTQGVKSLLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDDPPLLTRYLRIHFPQSWHQAIRMEVLGCEAQQDIY 2351
DB 2281 KVFQGNQDSFTPVVNSLDDPPLLTRYLRIHFPQSWHQAIRMEVLGCEAQQDIY 2332

RESULT 7

US-08-251-937A-4
; Sequence 4, Application: US/08251937A
; Patent No. 5583209
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Rudge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patcin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/251,937A
; FILING DATE: 31-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6367
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: protein	Query Match	98.9%;	Score 12282;	DB 1;	Length 2332;
HYPOTHETICAL: NO	Best Local Similarity	99.5%;	Pred. No. 0;		
ANTI-SENSE: NO	Matches 2327;	Conservative	2;	Mismatches	3;
FRAGMENT TYPE: N-terminal				Indels	0;
ORIGINAL SOURCE:					
ORGANISM: Homo sapien					
TISSUE TYPE: Liver cDNA sequence					
US-08-251-937A-4					
20	ATRRYYLGAVELSWDYMQSDLGELPVDARFPFRVPKSPFNNTSVYYKTLFVEFTDHLFN	79			
Db	1	ATRRYYLGAVELSWDYMQSDLGELPVDARFPFRVPKSPFNNTSVYYKTLFVEFTVHLEN	60		
QY	80	IAKPRPFWNGLLGPITQAEVYTVITLKNASHPVSLHAGVSWKASGABVDDQTSQ	139		
Db	61	IAKPRPFWNGLLGPITQAEVYTVITLKNASHPVSLHAGVSWKASGABVDDQTSQ	120		
QY	140	REKEDDKVPFGGSHTVVMQVLKENGPMASDPLCLTYISVLSDVKDLSGLIGALLVCR	199		
Db	121	REKEDDKVPFGGSHTVVMQVLKENGPMASDPLCLTYISVLSDVKDLSGLIGALLVCR	180		
QY	200	EGSLAKEKTQTLLKFTILLPAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYNR	259		
Db	181	EGSLAKEKTQTLLKFTILLPAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYNR	240		
QY	260	SLPGLICGRKSVYHVIQMGTTPEVHSIFLEGHTFLVRNHRQASLETSPIFTTAQTLL	319		
Db	241	SUPGLIGCRKSVYHVIQMGTTPEVHSIFLEGHTFLVRNHRQASLETSPIFTTAQTLL	300		
QY	320	MDLQGFLLSCHSSHOHDGMEAYVKVDSCPEPPOLIMKNBEAEADYDDDLTDSMDVYRF	379		
Db	301	MDLQGFLLSCHSSHOHDGMEAYVKVDSCPEPPOLIMKNBEAEADYDDDLTDSMDVYRF	360		
QY	380	DDNSPSFTQIRSAVKHKPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSYLNNGPQIG	439		
Db	361	DDNSPSFTQIRSAVKHKPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSYLNNGPQIG	420		
QY	440	RKYKVRFWAYTDEFTKTRTAEQHSGLIGPLLYGEVSDTLITLTKQASRPYNYPHGI	499		
Db	421	RKYKVRFWAYTDEFTKTRTAEQHSGLIGPLLYGEVSDTLITLTKQASRPYNYPHGI	480		
QY	500	TDVRPLYSRRLPGVKVHKLDPIELGELFKYKWTVTVDGPTKSDPRCLTRVYSSFYNME	559		
Db	481	TDVRPLYSRRLPGVKVHKLDPIELGELFKYKWTVTVDGPTKSDPRCLTRVYSSFYNME	540		
QY	560	RLASGLIGPLLYCYKESYDQGNQIMSDKRNVLFSYFDENRSWYLTENIQRELPNPAG	619		
Db	541	RLASGLIGPLLYCYKESYDQGNQIMSDKRNVLFSYFDENRSWYLTENIQRELPNPAG	600		
QY	620	VOLEDPFCASIMHSINGYVFDLSQLSVCLHEVAYWILSIGACTDPLSVFFSGYTFKH	679		
Db	601	VOLEDPFCASIMHSINGYVFDLSQLSVCLHEVAYWILSIGACTDPLSVFFSGYTFKH	660		
QY	680	KWVYEDTLTLFPFSGETVMSMENPGLMTLIGCHNSDFERNRGWTALLKYSSCDKNTGDYIE	739		
Db	661	KWVYEDTLTLFPFSGETVMSMENPGLMTLIGCHNSDFERNRGWTALLKYSSCDKNTGDYIE	720		
QY	740	DSYEDISAYLLSKNNAIEPRSFQNSRHSFTQKQFNATTIPENDIEKTDWFAHRTMP	799		
Db	721	DSYEDISAYLLSKNNAIEPRSFQNSRHSFTQKQFNATTIPENDIEKTDWFAHRTMP	780		
QY	800	KIQNVSSSDLLMLLRQSPFPHGLSLSDQAEKYEYTFSDPSPGALDSNNLSSEWTHFRPQ	859		
Db	781	KIQNVSSSDLLMLLRQSPFPHGLSLSDQAEKYEYTFSDPSPGALDSNNLSSEWTHFRPQ	840		
QY	860	LHSGDMVFTPESGLQRLNEXLGTATAELKCLDPKYSSTSNLLISTIPSDNLAAGTDN	919		
Db	841	LHSGDMVFTPESGLQRLNEXLGTATAELKCLDPKYSSTSNLLISTIPSDNLAAGTDN	900		

1981 GVFETVEMLPKXAGINWVECLIGSHLHAGNSTLFLVVSNNKQCPPLGWASGHIRDFOITAS 2040
2060 GOYGOWAPKILARLHYSGSINAWTKPFFSWIKVDLLAPMIIHGKIKTOGAROKFSSLYISQ 2119
2041 GOYGOWAPKILARLHYSGSINAWTKPFFSWIKVDLLAPMIIHGKIKTOGAROKFSSLYISQ 2100
2120 FIIMYSIDGKKWOTYRGNSTGTLWVFFGNVDSGIGKINENPPIIARYIRLHPTHYSIRS 2179
2101 FIIMYSIDGKKWOTYRGNSTGTLWVFFGNVDSGIGKINENPPIIARYIRLHPTHYSIRS 2160
2180 TLMELMGCDLNSCMLPLGMSKAISDAQITASSYFTNMATWSPSKARLHLQGRSNMWR 2239
2161 TLMELMGCDLNSCMLPLGMSKAISDAQITASSYFTNMATWSPSKARLHLQGRSNMWR 2220
2240 PQVNNKREWLOVDFQKTMKVTGVTQGVKSLLSMYVKEFLISSODGCHQWTLFFQNGKV 2299
2221 PQVNNKREWLOVDFQKTMKVTGVTQGVKSLLSMYVKEFLISSODGCHQWTLFFQNGKV 2280
2300 KVFQGNQDSFTPVVNSLDPPLLTRYLRHQPQSWHQAIAEVEVLGCEAODLY 2351
2281 KVFQGNQDSFTPVVNSLDPPLLTRYLRHQPQSWHQAIAEVEVLGCEAODLY 2332

RESULT 8

US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,133A
; FILING DATE: March 11, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU/76677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-572-6508
; TELEFAX: 404-572-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver CDNA sequence
US-08-212-133A-2

Query Match 98.9%; Score 12282; DB 1; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRRYYLGAVELSWDQMOSDLGELPVDARPPRPVPSKSPFNISVVKYKTLFVEFTDHLFN 79
DB 1 ATRRYYLGAVELSWDQMOSDLGELPVDARPPRPVPSKSPFNISVVKYKTLFVEFTDHLFN 60
QY 80 IAKPRPPWGLLGPPTQAEVYDTWVITLKNMASHPVSLHAGVSVYKASGAEYDDQTSQ 139
DB 61 IAKPRPPWGLLGPPTQAEVYDTWVITLKNMASHPVSLHAGVSVYKASGAEYDDQTSQ 120
QY 140 REKEDDKVPFGSGHTVWVOVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVPFGSGHTVWVOVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKQHTVNGVNR 259
DB 181 EGS LAKEKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKQHTVNGVNR 240
QY 260 SLPGLIGCHRKSVYVHVIENGITPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319
DB 241 SLPGLIGCHRKSVYVHVIENGITPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
QY 320 MDLGQFLLSCHSSHQHDGMEAYVKYVDSCEPEEPOLIMKNEEAEDYDDDLTDSMDVVRP 379
DB 301 MDLGQFLLSCHSSHQHDGMEAYVKYVDSCEPEEPOLIMKNEEAEDYDDDLTDSMDVVRP 360
QY 380 DDNSPSFQIRSVAKKPKTWVHYIAAFEDWDYAPLVAPDDRYSKSYLNNQFORIG 439
DB 361 DDNSPSFQIRSVAKKPKTWVHYIAAFEDWDYAPLVAPDDRYSKSYLNNQFORIG 420
QY 440 RYKVKRFWAYTDETFKTRBAIOHESGILGPLYGEVGDFTLLIFKNQASRPNIYPHGI 499
DB 421 RYKVKRFWAYTDETFKTRBAIOHESGILGPLYGEVGDFTLLIFKNQASRPNIYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEGFTKSDPRCLTRYSSFVME 559
DB 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEGFTKSDPRCLTRYSSFVME 540
QY 560 RDLASGLIGPLLI CYKESVDQKGNQIMSKRNVLFSVFDENRSWYITENIQRFLENPAG 619
DB 541 RDLASGLIGPLLI CYKESVDQKGNQIMSKRNVLFSVFDENRSWYITENIQRFLENPAG 600
QY 620 VOLDEPFOASNTMHSINGVVFDSLQLSVCLHVAWYIILSICAOQDFLSVFFSGYTFKH 679
DB 601 VOLDEPFOASNTMHSINGVVFDSLQLSVCLHVAWYIILSICAOQDFLSVFFSGYTFKH 660
QY 680 KMVYEDTLTLFPFSGBTVPFMSMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYE 739
DB 661 KMVYEDTLTLFPFSGBTVPFMSMENPGLWILGCHNSDFRNGMTALLKVSSCDKNTGDYE 720
QY 740 DSYEDISAYLLSKNNAIEPRSPNSRHSHTROKQFNATTIPENDIEKTDPEWHAHTMP 799
DB 721 DSYEDISAYLLSKNNAIEPRSPNSRHSHTROKQFNATTIPENDIEKTDPEWHAHTMP 780
QY 800 KIQNVSSDILLMLLRQSPTHGSLSDLOBAKYETFDSPSPCAIDSNNSLSMTFRPQ 859
DB 781 KIQNVSSDILLMLLRQSPTHGSLSDLOBAKYETFDSPSPCAIDSNNSLSMTFRPQ 840
QY 860 LHS GDMVFTPESGLOLRINEKLGTTAATLKKLDFKVSSTSNLLSTIPSDNLAAGTDN 919
DB 841 LHS GDMVFTPESGLOLRINEKLGTTAATLKKLDFKVSSTSNLLSTIPSDNLAAGTDN 900
QY 920 TSSIGPSPMPVHYDSQIDTTLFGKKSPPLTESGGPLSLSEENNDKLLBSGLNNSQESSW 979
DB 901 TSSIGPSPMPVHYDSQIDTTLFGKKSPPLTESGGPLSLSEENNDKLLBSGLNNSQESSW 960
QY 980 GKNVSTESGRLPKGRAGPALLTKDNALFKVSIILLTKNTSNKSNATNRKTHIDGPSL 1039
DB 961 GKNVSTESGRLPKGRAGPALLTKDNALFKVSIILLTKNTSNKSNATNRKTHIDGPSL 1020

QY 1040 LIENSPVQWQILSDTEFKKVTPLIHDRMLMDXNATRLNHNMSNKTSSKNMVMQOK 1039
DB 1021 LIENSPVQWQILSDTEFKKVTPLIHDRMLMDXNATRLNHNMSNKTSSKNMVMQOK 1030
QY 1100 KEGIPDDAQNPDMSFFKMLFLPSARWIQTHGKNSLNSCGGSPKOLVSLGPKSVG 1159
DB 1081 KEGIPDDAQNPDMSFFKMLFLPSARWIQTHGKNSLNSCGGSPKOLVSLGPKSVG 1140
QY 1160 ONFUSEKNKVVGGEFTKDVGLKEMVPPSRNPLTNLDNLHNNTHNQSKKIQBEIEK 1219
DB 1141 ONFUSEKNKVVGGEFTKDVGLKEMVPPSRNPLTNLDNLHNNTHNQSKKIQBEIEK 1200
QY 1220 KETLIQENNVLPQIRTVGTGTNFMKNLFLSTRQNVESYDCAVAPLQDPRSLNDSNR 1279
DB 1201 KETLIQENNVLPQIRTVGTGTNFMKNLFLSTRQNVESYDCAVAPLQDPRSLNDSNR 1260
QY 1280 TKKTAHFSKKEBENLEGLNQKQIYEKACTTRISPTNSQNFVQORSKRALQKRL 1339
DB 1261 TKKTAHFSKKEBENLEGLNQKQIYEKACTTRISPTNSQNFVQORSKRALQKRL 1320
QY 1340 PLETELEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEKGALTQSPLSDCLTRSHSI 1399
DB 1321 PLETELEKRIIVDDTSTQWKNMKHLTPSTLTQIDYNEKEKGALTQSPLSDCLTRSHSI 1380
QY 1400 POANRSPPIAKVSSFPRIPIYLRVLFDQNSSHLPAASYRKDQSGVOESSHFLQGA 1459
DB 1381 POANRSPPIAKVSSFPRIPIYLRVLFDQNSSHLPAASYRKDQSGVOESSHFLQGA 1440
QY 1460 NNLSLAULTLEMTGDORVSGLSGTSATNSVYKKVENTVLKPDLPKTSKGVKELLPKVHI 1519
DB 1441 NNLSLAULTLEMTGDORVSGLSGTSATNSVYKKVENTVLKPDLPKTSKGVKELLPKVHI 1500
QY 1520 YQKOLFPPTETNSGSHLDLVEGSLQTEGAIKWNEANRPGKPFPLRVATESSAKTSPK 1579
DB 1501 YQKOLFPPTETNSGSHLDLVEGSLQTEGAIKWNEANRPGKPFPLRVATESSAKTSPK 1560
QY 1580 LLDPLAWDNHYGTQIPKEBWSQKSPKTAFAKKDITLSLNACSNHAIAINEGQNK 1639
DB 1561 LLDPLAWDNHYGTQIPKEBWSQKSPKTAFAKKDITLSLNACSNHAIAINEGQNK 1620
QY 1640 ELEVWAKQGRTERLCSONPVPVLEKQREIRFTTLQSDQEEIDYDDTISVEMKEDFDIY 1699
DB 1621 ELEVWAKQGRTERLCSONPVPVLEKQREIRFTTLQSDQEEIDYDDTISVEMKEDFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSGSVQPKKVPQEBTD 1759
DB 1681 DEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNRAQSGSVQPKKVPQEBTD 1740
QY 1760 GSFTQPLVYRGELNEHGLLGPYIRAEVDNIMVTRNQASRPYSYSLISYEEBQROGA 1819
DB 1741 GSFTQPLVYRGELNEHGLLGPYIRAEVDNIMVTRNQASRPYSYSLISYEEBQROGA 1800
QY 1820 BPRKNFVKNETKTVFKVOHMAPTKDFDCKAWAYPSDVLKDVHSLGLPLLVCHT 1879
DB 1801 BPRKNFVKNETKTVFKVOHMAPTKDFDCKAWAYPSDVLKDVHSLGLPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQBFALPFTTIDETKSWYFTENMERNACPCNIQMEDPTPKENYRPHA 1939
DB 1861 NTLNPAHGRQVTVQBFALPFTTIDETKSWYFTENMERNACPCNIQMEDPTPKENYRPHA 1920
QY 1940 INGYIMDTPLGLWMAQDQIRIYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKALYNLYP 1999
DB 1921 INGYIMDTPLGLWMAQDQIRIYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKALYNLYP 1980
QY 2000 GYFETVEMLPKAGIWRVCLTGEHLHAGMSTLFLVYSNKCOTPLQMASGHIRDFOITAS 2059
DB 1981 GYFETVEMLPKAGIWRVCLTGEHLHAGMSTLFLVYSNKCOTPLQMASGHIRDFOITAS 2040
QY 2060 GOYGOWAPXKLARIHYSGSSINAWSTKEPFPFWIKVYDILLAPMIIGHIKTQGARQKFPSSLIYISQ 2119
DB 2041 GOYGOWAPXKLARIHYSGSSINAWSTKEPFPFWIKVYDILLAPMIIGHIKTQGARQKFPSSLIYISQ 2100
QY 2120 FTIYSLDGKKWQTYRGNSTGTLMVFGNVDSSGIKHNIFNPPIARIYRLHPHYSIRS 2179

DB 2101 FTIYSLDGKKWQTYRGNSTGTLMVFGNVDSSGIKHNIFNPPIARIYRLHPHYSIRS 2160
QY 2180 TLRMELMGCDLNSCMLGWSKAISSAQITASSYFTNMFEATWSPSKARLHLQGRSNAWR 2239
DB 2161 TLRMELMGCDLNSCMLGWSKAISSAQITASSYFTNMFEATWSPSKARLHLQGRSNAWR 2220
QY 2240 PQVNNPKREWQVDPQKTMKVTGTGTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 PQVNNPKREWQVDPQKTMKVTGTGTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KVPQGNQDSPTPVVNSLDDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 2351
DB 2281 KVPQGNQDSPTPVVNSLDDPPLLTRYLRIHPQSWHQAIALRMEVLGCEAODLY 2332

RESULT 9
US-08-474-503-2
; Sequence 2, Application US/08474503
; Patent No. 574446
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,503
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMDU06CIP(3)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6500
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
US-08-474-503-2

Query Match 96.9%; Score 12282; DB 1; Length 2332;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 20 ATRRYLGAVELSDYMQSDLGELPVDARFPFRPKSPPTNTSVYKKTILVEFTDHLFN 79
DB 1 ATRRYLGAVELSDYMQSDLGELPVDARFPFRPKSPPTNTSVYKKTILVEFTDHLFN 60
QY 80 IAKPREPPMGLLGTIOAEVYDVTITLKNVASHPVSLHVGVSYWKASGAEYDDQTSQ 139
DB 61 IAKPREPPMGLLGTIOAEVYDVTITLKNVASHPVSLHVGVSYWKASGAEYDDQTSQ 120

140 REKEDDXVPGGSHYYVWVLEKNGPMASDPLCLTYSYLSHVLDLVKDLNSGLGALLVCR 139
Db 121 REKEDDXVPGGSHYYVWVLEKNGPMASDPLCLTYSYLSHVLDLVKDLNSGLGALLVCR 180
Qy 200 EGSIAKEKTOTLHKFILLFAVDEGKSHWSETKNSLMQDRDAASARAWPKXHTVNGYVNR 259
Db 181 EGSIAKEKTOTLHKFILLFAVDEGKSHWSETKNSLMQDRDAASARAWPKXHTVNGYVNR 240
Qy 260 SLPLGLICHRKSVVWVIGVGTTPVEHSIFLEBHTFLVRNHRQASLEISPIITLTAQTL 319
Db 241 SLPLGLICHRKSVVWVIGVGTTPVEHSIFLEBHTFLVRNHRQASLEISPIITLTAQTL 300
Qy 320 MDLGOFLLSCHISSHQHDGMEAYVYKVDSCPEPOLINKNNEEAEDYDDDLTDSMDVVR 379
Db 301 MDLGOFLLSCHISSHQHDGMEAYVYKVDSCPEPOLINKNNEEAEDYDDDLTDSMDVVR 360
Qy 380 DDNSPSFQIRSVAKKHPTVWVHYIAAEEEDMDYAPLVAPDDRYSKYSLNNGPQIRG 439
Db 361 DDNSPSFQIRSVAKKHPTVWVHYIAAEEEDMDYAPLVAPDDRYSKYSLNNGPQIRG 420
Qy 440 RYKYZKRFMAYTDTFTKTRAIQHESGILGPLLVEVGDILLIIIFKXQASRPYNIYPHGI 499
Db 421 RYKYZKRFMAYTDTFTKTRAIQHESGILGPLLVEVGDILLIIIFKXQASRPYNIYPHGI 480
Qy 500 TDVREPLYSRLLPKGVKHLKOPPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 559
Db 481 TDVREPLYSRLLPKGVKHLKOPPLPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSPVNME 540
Qy 560 RDLASGLIGPLLI CYKESVDQKGNQIMSDKRNVLFSVDFENRSMWLTENIQRELNPAG 619
Db 541 RDLASGLIGPLLI CYKESVDQKGNQIMSDKRNVLFSVDFENRSMWLTENIQRELNPAG 600
Qy 620 VQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYVYLLSIGAQOTPLSVFFSGYTFKH 679
Db 601 VQLEDPEFQASNIHMSINGYVFDLSQLSVCLHEVAYVYLLSIGAQOTPLSVFFSGYTFKH 660
Qy 680 KMVYEDTLTLPFSGEIVFMSMENPGIWLICHNSDFRNRMNTALLVSSCDKXNGDYVE 739
Db 661 KMVYEDTLTLPFSGEIVFMSMENPGIWLICHNSDFRNRMNTALLVSSCDKXNGDYVE 720
Qy 740 DSYEDI SAYLLSKXNAI BPSFSQNSRHPSTRQKQFNATTIPENDIEKTDWPAHRTMP 799
Db 721 DSYEDI SAYLLSKXNAI BPSFSQNSRHPSTRQKQFNATTIPENDIEKTDWPAHRTMP 780
Qy 800 KIQVSSSDLLMLRQSPTRPCLSLSDLOERAKYTFSDDPSPGALDSNNSLSETHFRPQ 859
Db 781 KIQVSSSDLLMLRQSPTRPCLSLSDLOERAKYTFSDDPSPGALDSNNSLSETHFRPQ 840
Qy 860 LHHSGDMVFT PESGLQRLNKLGTAAATELKLDFKVSSTSNLI STIPSDNLAAGTDN 919
Db 841 LHHSGDMVFT PESGLQRLNKLGTAAATELKLDFKVSSTSNLI STIPSDNLAAGTDN 900
Qy 920 TSSILGPPSMFVHYDSQDITLLFGKKSPLTBSGGPLSISENNDSKLLLESGLMNSQBSW 979
Db 901 TSSILGPPSMFVHYDSQDITLLFGKKSPLTBSGGPLSISENNDSKLLLESGLMNSQBSW 960
Qy 980 GKNVSSSTESGLFKGKAHGPAJLTKONALFKVSI SLLKTKNTSNNSATNRKTHIDGPSL 1039
Db 961 GKNVSSSTESGLFKGKAHGPAJLTKONALFKVSI SLLKTKNTSNNSATNRKTHIDGPSL 1020
Qy 1040 LIENSPPSVWQNLBSDETFKFKVTPLIHDMRLMDKNATALRNHMSNKTTSKKNMEMVQOK 1099
Db 1021 LIENSPPSVWQNLBSDETFKFKVTPLIHDMRLMDKNATALRNHMSNKTTSKKNMEMVQOK 1080
Qy 1100 KEGPIPPDAQNPDMSPFKMLFLPESARWIORTHGKNSLNSQGPSPKQVLVSLGPEKSVEG 1159
Db 1081 KEGPIPPDAQNPDMSPFKMLFLPESARWIORTHGKNSLNSQGPSPKQVLVSLGPEKSVEG 1140
Qy 1160 QNFILSEKNKVVVKGCEFTKQVGLKEMVTPSSENFLTNLDNLHNNHTNQBKKIQEBIEK 1219
Db 1141 QNFILSEKNKVVVKGCEFTKQVGLKEMVTPSSENFLTNLDNLHNNHTNQBKKIQEBIEK 1200

1220 KETLIOENVVLPOIHTVITGTKNPMKNLFLLSSTRQNVESGYDGAYAPVLODFRSLNDSTNR 1279
Db 1201 KETLIOENVVLPOIHTVITGTKNPMKNLFLLSSTRQNVESGYDGAYAPVLODFRSLNDSTNR 1260
Qy 1280 TKKHTAHFKKGBEENLEGLGNQTKOIVKRYACTTRISNPTSQONFVTOBSKQALQFRL 1339
Db 1261 TKKHTAHFKKGBEENLEGLGNQTKOIVKRYACTTRISNPTSQONFVTOBSKQALQFRL 1320
Qy 1340 PLEETELEKXIIIVDDTSTQWSKNMKHLTPSTLFOIDYNEKEKGAITQSPUSDCLTSHSI 1399
Db 1321 PLEETELEKXIIIVDDTSTQWSKNMKHLTPSTLFOIDYNEKEKGAITQSPUSDCLTSHSI 1380
Qy 1400 PQANRSPPLPIAKVSSPSPRIPIYLTIVLFDONSSHLPAAASYRKDKSGVDESSHFLOAKK 1459
Db 1381 PQANRSPPLPIAKVSSPSPRIPIYLTIVLFDONSSHLPAAASYRKDKSGVDESSHFLOAKK 1440
Qy 1460 NNLISLAILTLEMTGDQREVSGLSGT SATNSVYTKKVENTVL PKDPLPKTSKVKELLPKVHI 1519
Db 1441 NNLISLAILTLEMTGDQREVSGLSGT SATNSVYTKKVENTVL PKDPLPKTSKVKELLPKVHI 1500
Qy 1520 YOXDLPTETSNGPSGHLDLVEGSLLOGTREGAIKWNENRPGKVPFLRVATESAKTPSK 1579
Db 1501 YOXDLPTETSNGPSGHLDLVEGSLLOGTREGAIKWNENRPGKVPFLRVATESAKTPSK 1560
Qy 1580 LLDPLAWMDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTILSLNACESHAAIAINEGONKP 1639
Db 1561 LLDPLAWMDNHYGTQIPKEBWKSOEKSPEKTAFFKKDTILSLNACESHAAIAINEGONKP 1620
Qy 1640 EIEVTWAKQORTERLCSQNPVVKRHOEITRTTLOSQDQBEIDYDDTISVEMKEDDIY 1699
Db 1621 EIEVTWAKQORTERLCSQNPVVKRHOEITRTTLOSQDQBEIDYDDTISVEMKEDDIY 1680
Qy 1700 DEBENQSPRSFOKTRHYFAAVERLWDYGMSSSPHVLNRASGVSVPQPKVXVFOFTD 1759
Db 1681 DEBENQSPRSFOKTRHYFAAVERLWDYGMSSSPHVLNRASGVSVPQPKVXVFOFTD 1740
Qy 1760 GSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEDDQOGA 1819
Db 1741 GSFTQPLRYGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEDDQOGA 1800
Qy 1820 EPRKNFKVKNETKTYFWKVOHHMAPTKDEDFCKAWAYFSDVDLEKDVHSLIGPLLCHT 1879
Db 1801 EPRKNFKVKNETKTYFWKVOHHMAPTKDEDFCKAWAYFSDVDLEKDVHSLIGPLLCHT 1860
Qy 1880 NTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMBNCRAPCNIOEMEDTFKENYRPHA 1939
Db 1861 NTLNPAHGRQVTVQEPALFTTIFDETKSWYFTENMBNCRAPCNIOEMEDTFKENYRPHA 1920
Qy 1940 INGVIMDTLEGLYMAQDQRIRWYLLSMGSNENIHSIFSGHVFTVRKKEBYKXALYNLYP 1999
Db 1921 INGVIMDTLEGLYMAQDQRIRWYLLSMGSNENIHSIFSGHVFTVRKKEBYKXALYNLYP 1980
Qy 2000 GVPETVEMLSKAGIWRVECLIGEHLAGHAGSTLFLVYSNKCOTPLGNWASGHIRDFQITAS 2059
Db 1981 GVPETVEMLSKAGIWRVECLIGEHLAGHAGSTLFLVYSNKCOTPLGNWASGHIRDFQITAS 2040
Qy 2060 GOYQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGIKTOGARQKFSLSYISQ 2119
Db 2041 GOYQOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIIHGIKTOGARQKFSLSYISQ 2100
Qy 2120 FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGICKINIPNPPIIARYIRLHPHYHYSIRS 2179
Db 2101 FIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGICKINIPNPPIIARYIRLHPHYHYSIRS 2160
Qy 2180 TLRMELMGCDLNSCNPGLMESKAI SDAQITASSYFTNVEATWSPSKARHLQGSNAWR 2239
Db 2161 TLRMELMGCDLNSCNPGLMESKAI SDAQITASSYFTNVEATWSPSKARHLQGSNAWR 2220
Qy 2240 PQVNNPKEWLVQDPQKTMKYTGVTTCQVKSLLSMYVKEFLISSQDGHQWTLFFQNGKV 2299
Db 2221 PQVNNPKEWLVQDPQKTMKYTGVTTCQVKSLLSMYVKEFLISSQDGHQWTLFFQNGKV 2280
Qy 2300 KVFQGNQDSFTFVVNSLDPPLLTRYLRIHFQSWVHQAIRLXREVLGCEAQDLY 2351

2281 KVFQGNQDSFTEPVNSLDPELLTRYLRHPQSVWVHQIALRMEVLGCEAQDLY 2332

```
Db 121 REKEDDKVFFGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLGALLVCR 180
QY 200 EGSIAKEKTTQTLKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKWHTVNGVNR 259
Db 181 EGSIAKEKTTQTLKFILLFAVDFEGKSWHSETKNSLMQDRDAASARAWPKWHTVNGVNR 240
QY 260 SLPLGICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSIFITLTAQTL 319
Db 241 SLPLGICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHRQASLSIFITLTAQTL 300
QY 320 MDLCQFLLSCHISSHQHDCMEAYKVDSCPEEPOLIMKNNEEADYDDDLTDSMDVVRP 379
Db 301 MDLCQFLLFCHISSHQHDCMEAYKVDSCPEEPOLRMKNNEEADYDDDLTDSMDVVRP 360
QY 380 DDDNSPFIQIRSVAKKPKTWVHYIAAEBEDWDYAPLVLAPDDRSYKSOYLNNGPQIRG 439
Db 361 DDDNSPFIQIRSVAKKPKTWVHYIAAEBEDWDYAPLVLAPDDRSYKSOYLNNGPQIRG 420
QY 440 RYKVKVRPMAYTDETFKTRBAIOHESGILGPLLYGEGVDTLLIIFKQASPPYNIYPHGI 499
Db 421 RYKVKVRPMAYTDETFKTRBAIOHESGILGPLLYGEGVDTLLIIFKQASPPYNIYPHGI 480
QY 500 TDVRLYSRRLPKGVKHLKOPPLPGRIEYKVKWTVVEDGPTKSDPRCLTRYVSSPVNME 559
Db 481 TDVRLYSRRLPKGVKHLKOPPLPGRIEYKVKWTVVEDGPTKSDPRCLTRYVSSPVNME 540
QY 560 RDLASGLTGILLICYKESVDQKQIMSDKENVILFSVFDENRNSWYLPENIQRLEPNPAG 619
Db 541 RDLASGLTGILLICYKESVDQKQIMSDKENVILFSVFDENRNSWYLPENIQRLEPNPAG 600
QY 620 VQLEDPEFOASNIHNSINGYVFDLSQLSVCLHVEVAYWYILSIGAQTDPLSFVFGYFKH 679
Db 601 VQLEDPEFOASNIHNSINGYVFDLSQLSVCLHVEVAYWYILSIGAQTDPLSFVFGYFKH 660
QY 680 KMVYEDTLTLFPFSGETVFMSENPLGLMILGCHNSDPNRGWTALLKVSSCDKNTGDIYE 739
Db 661 KMVYEDTLTLFPFSGETVFMSENPLGLMILGCHNSDPNRGWTALLKVSSCDKNTGDIYE 720
QY 740 DSYEDISAYLLSKXNAIEPRSFSONSRHPTROQFNATTIPENDIEKTDWFAHRTMP 799
Db 721 DSYEDISAYLLSKXNAIEPRSFSONSRHPTROQFNATTIPENDIEKTDWFAHRTMP 780
QY 800 KIQVSSSDLLMLAROSPTPHGLSLSLQEAKEYTFSDPSGALDSNNSSEWTHPRPQ 859
Db 781 KIQVSSSDLLMLAROSPTPHGLSLSLQEAKEYTFSDPSGALDSNNSSEWTHPRPQ 840
QY 860 LHSXDMVFTPESGLOLRNEKLGTTAATELKLKLDPKVSSSTSNLIPIPSDNLAAGTDN 919
Db 841 LHSXDMVFTPESGLOLRNEKLGTTAATELKLKLDPKVSSSTSNLIPIPSDNLAAGTDN 900
QY 920 TSSLGPPSNPVHYDQSLDTTLFGKXSSPLTSGGPLSLSENNNSKLLSGLMNSQESSW 979
Db 901 TSSLGPPSNPVHYDQSLDTTLFGKXSSPLTSGGPLSLSENNNSKLLSGLMNSQESSW 960
QY 980 GKXVSSSTSGRLFKGKRAHGPAALLTKONALFKVSIISLKTNTKSNNSATNKTHIDGSL 1039
Db 961 GKXVSSSTSGRLFKGKRAHGPAALLTKONALFKVSIISLKTNTKSNNSATNKTHIDGSL 1020
QY 1040 LIENSFSPVWQNILESDETFKVTPLIHDRMLMDKNATALRNLHNSNKTTSKKNEMVQOK 1099
Db 1021 LIENSFSPVWQNILESDETFKVTPLIHDRMLMDKNATALRNLHNSNKTTSKKNEMVQOK 1080
QY 1100 KEGPIPPDAQNPDMSEFFKMLFLPESARWIQTHGHKNSLNSCGGSPKQVLGLGPEKSVEG 1159
Db 1081 KEGPIPPDAQNPDMSEFFKMLFLPESARWIQTHGHKNSLNSCGGSPKQVLGLGPEKSVEG 1140
QY 1160 QNFLEKKNKVVKGFEFTKQVCLKEMVFPSSRLPLTNLDLHNHNTNOKKIQEETEK 1219
Db 1141 QNFLEKKNKVVKGFEFTKQVCLKEMVFPSSRLPLTNLDLHNHNTNOKKIQEETEK 1200
QY 1220 KETLQENNVLPQIHVTGTXNFMKNLFLLSPTROWVEGSDYGAAPVLQDPRSLNDSNR 1279
Db 1201 KETLQENNVLPQIHVTGTXNFMKNLFLLSPTROWVEGSDYGAAPVLQDPRSLNDSNR 1260
```

2281 KVFQGNQDSFTEPVNSLDPELLTRYLRHPQSVWVHQIALRMEVLGCEAQDLY 2332

RESULT 10
US-08-670-707A-2
Sequence 2, Application US/08670707A
Patent No. 5859204
GENERAL INFORMATION:

APPLICANT: Iollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTX: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,707A
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:

NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95F

TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: YES

ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Liver

US-08-670-707A-2
Query Match 98.9%; Score 12282; DB 2; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 20 ATRRYVGLAVELSWDMQSDLGELPVDARPPRPVPSFPFN-SVYVYKTLFVEFTDHLFN 7s

Db 1 ATRRYVGLAVELSWDMQSDLGELPVDARPPRPVPSFPFN-SVYVYKTLFVEFTVHLFN 60

QY 80 IAKPRPPWMLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYKASGEGAEYDDQTSQ 139

Db 61 IAKPRPPWMLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYKASGEGAEYDDQTSQ 120

QY 140 REKEDDKVFFGGSHTYVWQVLKENGPMASDPLCLITYSLSHVDLVKDLNGLGALLVCR 199

QY 1280 TKCHTAHFSKGBBENLEGLNQTKQIIVEKYACTTRISPNSTQONFVQTSKALKQFRL 1339
DB 1261 TKCHTAHFSKGBBENLEGLNQTKQIIVEKYACTTRISPNSTQONFVQTSKALKQFRL 1320
QY 1340 PLEETELEKRIIIVDDTSTQSKWKHUTPSTLTQIDYNEKEKGAITQSPSLSDCLTRHESI 1399
DB 1321 PLEETELEKRIIIVDDTSTQSKWKHUTPSTLTQIDYNEKEKGAITQSPSLSDCLTRHESI 1380
QY 1400 PQANRSLPIAKVSSFFSIRPIYILTRVLFDQNSHLPAASYRKDQSGVQSSHFLOQAKK 1459
DB 1381 PQANRSLPIAKVSSFFSIRPIYILTRVLFDQNSHLPAASYRKDQSGVQSSHFLOQAKK 1440
QY 1460 NNLSLAILTLEMTQDQREVQSLGTSATNSVTKKVENTVLPKPLPKTSGKVLELLPKVHI 1519
DB 1441 NNLSLAILTLEMTQDQREVQSLGTSATNSVTKKVENTVLPKPLPKTSGKVLELLPKVHI 1500
QY 1520 YQKDLFPETSNQSPGHLDLVEGSLLOQTEGAIKWEANRPGKVPFLRVATESAKTFSK 1579
DB 1501 YQKDLFPETSNQSPGHLDLVEGSLLOQTEGAIKWEANRPGKVPFLRVATESAKTFSK 1560
QY 1580 LLDPLAWNHYGTQIPKEWKSQBSPEKTAFFKDDTILSNACESHAAIAINEGQNK 1639
DB 1561 LLDPLAWNHYGTQIPKEWKSQBSPEKTAFFKDDTILSNACESHAAIAINEGQNK 1620
QY 1640 EIEVTKAGHTEBLCSQNPVLKSHOREITRTTLQSDQBEIDYDDTISVEMKXEDFDIY 1699
DB 1621 EIEVTKAGHTEBLCSQNPVLKSHOREITRTTLQSDQBEIDYDDTISVEMKXEDFDIY 1680
QY 1700 DEDNQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVQPKKVVFOEFTD 1759
DB 1681 DEDNQSPRSFQKTRHYFIAAVERLDWYGMSSSPHVLNRNAQSGSVQPKKVVFOEFTD 1740
QY 1760 GSFTQPIYRGELNHLGLLGYIRAEVNDNTMTVTRNQASRPYFSSLSIYBBDQOQGA 1819
DB 1741 GSFTQPIYRGELNHLGLLGYIRAEVNDNTMTVTRNQASRPYFSSLSIYBBDQOQGA 1800
QY 1820 EPRKNVZNETKYFWKVOHMAPTKDEPCKAWAYFSDVLEKDVHSGHIGLGLVCHT 1879
DB 1801 EPRKNVZNETKYFWKVOHMAPTKDEPCKAWAYFSDVLEKDVHSGHIGLGLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFFIFDETBSWFTENMERNCRAPCNQMEDPTFKENYRPHA 1939
DB 1861 NTLNPAHGRQVTVQEFALFFIFDETBSWFTENMERNCRAPCNQMEDPTFKENYRPHA 1920
QY 1940 INGVIMDTPLGLVMAQDQRIWYLLSMGSNENIHSIHPSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGVIMDTPLGLVMAQDQRIWYLLSMGSNENIHSIHPSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFEVEMLPKAGIWRVECLIGEHLHAGMTFLFVYSNKQCTPLGMASGHIRDFQITAS 2059
DB 1981 GVFEVEMLPKAGIWRVECLIGEHLHAGMTFLFVYSNKQCTPLGMASGHIRDFQITAS 2040
QY 2060 QYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLAPMIHIGIKTOGARQKFSLSYISQ 2119
DB 2041 QYQGWAPKLARLHYSGSINAWSTKEPFSWIKVDLAPMIHIGIKTOGARQKFSLSYISQ 2100
QY 2120 FIIMYSLDCKWQTVRGSTGTLVYFFGNVDSGGIKHNIFFPPIIARVIRLHPHYSIRS 2179
DB 2101 FIIMYSLDCKWQTVRGSTGTLVYFFGNVDSGGIKHNIFFPPIIARVIRLHPHYSIRS 2160
QY 2180 TLRMELMGCDLNSCSNPLQMSKALSDAQITASSYFTNNKATWSPSKARLHLQGSNAWR 2239
DB 2161 TLRMELMGCDLNSCSNPLQMSKALSDAQITASSYFTNNKATWSPSKARLHLQGSNAWR 2220
QY 2240 PQVNNPKSLQVDFQTKVTGVTQYKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
DB 2221 PQVNNPKSLQVDFQTKVTGVTQYKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KYFQGNQDSFTPVVNSLDFPPLITRYLRHQPQSWHQAIALRMEVLGCEAQDLY 2351
DB 2281 KYFQGNQDSFTPVVNSLDFPPLITRYLRHQPQSWHQAIALRMEVLGCEAQDLY 2332

RESULT 11
US-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
US-09-037-601-2

Query Match 98.9%; Score 12282; DB 3; Length 2332;
Best Local Similarity 99.8%; Pred. No 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRYYILGAVLSWDYMQSDIGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTDHLFN 79
DB 1 ATRYYILGAVLSWDYMQSDIGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTVHLFN 60
QY 80 IAKPRPPMGLGPTIQAEVYDTVITLKNMASHPVSLHVGSIYWKASGAEYDDQTSQ 139
DB 61 IAKPRPPMGLGPTIQAEVYDTVITLKNMASHPVSLHVGSIYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVFPQGSHTYVQVLKENGPMASDPLCLUTYSYLSHVDLVKDLNSGLIGALLVCR 199
DB 121 REKEDDKVFPQGSHTYVQVLKENGPMASDPLCLUTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 BGLAKEXTQTLLHKFILLFAVDFECKSWHSETKNSLMQDRDAASARAWPKMHTVNGYNR 259

Db 181 EGSIAKXNTQILHKEFLILFAVDEGKSEWETKNSLMOQDRDAASARAPKXHTVNGYVR 240
QY 260 SLPLGIGCHRSVYVHWITGMGTTPEVHISIFLEGHTFLVNRHQASLEISPTIFLTAQTLL 319
Db 241 SLPLGIGCHRSVYVHWITGMGTTPEVHISIFLEGHTFLVNRHQASLEISPTIFLTAQTLL 300
QY 320 MDLGOFLLSCHISSHCHDGMAYVAVDSCPEPOLIMKNNEAEEDYDDDLTDSMDVVR 379
Db 301 MDLGOFLLSCHISSHCHDGMAYVAVDSCPEPOLIMKNNEAEEDYDDDLTDSMDVVR 360
QY 380 DDNSPSEFIQRSVAKHPKTVWVHYIAAEEDWDYAPLVLPDDRSYKSYLNNQPORIG 439
Db 361 DDNSPSEFIQRSVAKHPKTVWVHYIAAEEDWDYAPLVLPDDRSYKSYLNNQPORIG 420
QY 440 RYKVKVRMAYVDETFKTRAEAOHESGILGPLLYGEVGTLLIFPKNOASRPYNIYPHGI 499
Db 421 RYKVKVRMAYVDETFKTRAEAOHESGILGPLLYGEVGTLLIFPKNOASRPYNIYPHGI 480
QY 500 TVDRPLYRRLPKGVKHLKDRPILPGEITPKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 559
Db 481 TVDRPLYRRLPKGVKHLKDRPILPGEITPKYKWTVTVEDGPTKSDPRCLTRYSSFVNM 540
QY 560 ZDLASGLICPLLCYKESVDQKNOIMSDKRVILFVSFVFNENRSLYLTENIQRFLENPAG 619
Db 541 ZDLASGLICPLLCYKESVDQKNOIMSDKRVILFVSFVFNENRSLYLTENIQRFLENPAG 600
QY 620 VOLEPPEQASIMHSINGVYFDSIQLSVCLHEVAYWILSTGAOTDPLSVFFSGYTKH 679
Db 601 VOLEPPEQASIMHSINGVYFDSIQLSVCLHEVAYWILSTGAOTDPLSVFFSGYTKH 660
QY 680 KMVYEDTTLTLPFSGETVMSMENPGLWILGCHNSDFNRNGMTALLKYSSCDKNTGDY 739
Db 661 KMVYEDTTLTLPFSGETVMSMENPGLWILGCHNSDFNRNGMTALLKYSSCDKNTGDY 720
QY 740 DSYEDISAYLLSKNAIBPRFSQNSRHPSRQKQFNATTIPENDIEKTDPMFAHRTMP 799
Db 721 DSYEDISAYLLSKNAIBPRFSQNSRHPSRQKQFNATTIPENDIEKTDPMFAHRTMP 780
QY 800 KIONVSSDLMMLRQSPPTGSLSDIOEAKYETPSDDPSGATDKNNSLSEWTHRPQ 859
Db 781 KIONVSSDLMMLRQSPPTGSLSDIOEAKYETPSDDPSGATDKNNSLSEWTHRPQ 840
QY 860 LHSQDMVFTPESGQLRINEKLGTTAATLKKLDFKVSSTSNLITIPSDNLAAGTDN 919
Db 841 LHSQDMVFTPESGQLRINEKLGTTAATLKKLDFKVSSTSNLITIPSDNLAAGTDN 900
QY 920 TSSIGPSPMZYHDSQDITLPGKXSSPLTESGGPLSISEENNDSKLLSGLMNSQESSW 979
Db 901 TSSIGPSPMZYHDSQDITLPGKXSSPLTESGGPLSISEENNDSKLLSGLMNSQESSW 960
QY 980 GKNVSTSGRLFKGRAGPALLTKONALFKVYSISLLTKNTKSNNSATNRKTHIDGPSL 1039
Db 961 GKNVSTSGRLFKGRAGPALLTKONALFKVYSISLLTKNTKSNNSATNRKTHIDGPSL 1020
QY 1040 LIENSFVWQNTLESTEFKVTPLDHNLMKDNATLRLNHNKNTTSSKNMVMVQK 1099
Db 1021 LIENSFVWQNTLESTEFKVTPLDHNLMKDNATLRLNHNKNTTSSKNMVMVQK 1080
QY 1100 KEGPIPPDACNPMDFKMLFLPESARWIORTHGKNSLNSGOGSPKQVLSIGPKSV 1159
Db 1081 KEGPIPPDACNPMDFKMLFLPESARWIORTHGKNSLNSGOGSPKQVLSIGPKSV 1140
QY 1160 QNFLSEKKNVYVKGKFTKVDGLKENVFPSSRNLFJTNLNHNKNTTSSKNMVMVQK 1219
Db 1141 QNFLSEKKNVYVKGKFTKVDGLKENVFPSSRNLFJTNLNHNKNTTSSKNMVMVQK 1200
QY 1220 KETLIQENVVLPOIHTVGTGNFMKNLFLISTRONVEGSDGAYAPVLOFRLNDSTNR 1279
Db 1201 KETLIQENVVLPOIHTVGTGNFMKNLFLISTRONVEGSDGAYAPVLOFRLNDSTNR 1260
QY 1280 TKKHTAFPSKKEBENLEGLNQTKQIWEKYACTTRISPNTSQNFVQTSKRALKOPRL 1339

RESULT 12
US-09-315-179-2

Db 1261 TKKHTAFPSKKEBENLEGLNQTKQIWEKYACTTRISPNTSQNFVQTSKRALKOPRL 1320
QY 1340 PLEETELEKRIIVDDTSTQSKNKHILTPSLTQIDYNEKEGALTOSPISDCUTRSHSI 1399
Db 1321 PLEETELEKRIIVDDTSTQSKNKHILTPSLTQIDYNEKEGALTOSPISDCUTRSHSI 1380
QY 1400 POANRSLPLIAKVSFFSIRPIYILTRVLFODNSSHLPAASRYKXKXSGVQESSHFLQAKK 1459
Db 1381 POANRSLPLIAKVSFFSIRPIYILTRVLFODNSSHLPAASRYKXKXSGVQESSHFLQAKK 1440
QY 1460 NNLISALITLTEMTOGOREVSLGTSATNSVTYKVENIVLPKPDLPKTSKGVKELLPKVHI 1519
Db 1441 NNLISALITLTEMTOGOREVSLGTSATNSVTYKVENIVLPKPDLPKTSKGVKELLPKVHI 1500
QY 1520 YOKDLFPTETISNGSPGHLDLVEGSLLOGTGCAIKWNEANRPKVPFLRVATESSAKTPSK 1579
Db 1501 YOKDLFPTETISNGSPGHLDLVEGSLLOGTGCAIKWNEANRPKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLANDNHYGTQIPKEEWKSQEKSPKTAFAKKDKTILSLNACSNHAIILNACSNHAI 1639
Db 1561 LLDPLANDNHYGTQIPKEEWKSQEKSPKTAFAKKDKTILSLNACSNHAIILNACSNHAI 1620
QY 1640 BIEVTWAKQGRTEBLCSQNPVVLKQHOBEITRITLQSDOBEIDYDDTISVEMKEDFDIY 1699
Db 1621 BIEVTWAKQGRTEBLCSQNPVVLKQHOBEITRITLQSDOBEIDYDDTISVEMKEDFDIY 1680
QY 1700 DEDENQSPRSFQKTRHYFIAAVERLWDYGSSSPHVLNRNAQSGSVPOPKVVFQBT 1759
Db 1681 DEDENQSPRSFQKTRHYFIAAVERLWDYGSSSPHVLNRNAQSGSVPOPKVVFQBT 1740
QY 1760 GSFTQPLVYRGLNHLGGLGYIRAEVEDNIMVTFRNQASRPYFYSLSIYEZDQROGA 1819
Db 1741 GSFTQPLVYRGLNHLGGLGYIRAEVEDNIMVTFRNQASRPYFYSLSIYEZDQROGA 1800
QY 1820 EPRKNFKPNETKTYFKWKVQHHMARTKDEPOCKAWAYFSDVDLEKDVHSGILGLVCHT 1879
Db 1801 EPRKNFKPNETKTYFKWKVQHHMARTKDEPOCKAWAYFSDVDLEKDVHSGILGLVCHT 1860
QY 1880 NTLNPAHQGVTVCEFFALFTTIFDETKSWYFTENWERNCRAPCNIQMEDPTFKENYRPHA 1939
Db 1861 NTLNPAHQGVTVCEFFALFTTIFDETKSWYFTENWERNCRAPCNIQMEDPTFKENYRPHA 1920
QY 1940 INGYIMDTLPGVNAQDQRIWYLLSMGSENHISIHPSGHVFTVRKKEBYKMALYNLYP 1999
Db 1921 INGYIMDTLPGVNAQDQRIWYLLSMGSENHISIHPSGHVFTVRKKEBYKMALYNLYP 1980
QY 2000 GVFEETVMLPSKAGIWRVECLIGELHAGMSTFLVYSNKCTOTPLGMAAGHIRDFOITAS 2059
Db 1981 GVFEETVMLPSKAGIWRVECLIGELHAGMSTFLVYSNKCTOTPLGMAAGHIRDFOITAS 2040
QY 2060 GOYGOWAPKLARLEYSGSINAWSTKEPFSWIKVDLLAPMIHIGIKTQARQKFSLSYSQ 2119
Db 2041 GOYGOWAPKLARLEYSGSINAWSTKEPFSWIKVDLLAPMIHIGIKTQARQKFSLSYSQ 2100
QY 2120 FIIMYSLDGKKWOYIRGNSGTGLMVFVGNVDSSGIKHNIHNPPIIARVIRLHPHYSIES 2179
Db 2101 FIIMYSLDGKKWOYIRGNSGTGLMVFVGNVDSSGIKHNIHNPPIIARVIRLHPHYSIES 2160
QY 2180 TLRMELMGCDLNSCNPGLMESKALSDAQITASSYFTNNFATWSPSKARLHLQGRSNWR 2239
Db 2161 TLRMELMGCDLNSCNPGLMESKALSDAQITASSYFTNNFATWSPSKARLHLQGRSNWR 2220
QY 2240 POWNPKEWLQVDFOKTMKVTVGTTTQGVKSLTSMYKKEPLISSQDQCHQWTLFPQNGKV 2299
Db 2221 POWNPKEWLQVDFOKTMKVTVGTTTQGVKSLTSMYKKEPLISSQDQCHQWTLFPQNGKV 2280
QY 2300 KVFQGNQDSFTPVNSLDPPILTRVIRTHPOSWHQIALRMEVLGCEAQDLY 2351
Db 2281 KVFQGNQDSFTPVNSLDPPILTRVIRTHPOSWHQIALRMEVLGCEAQDLY 2332

1681 DEDENQSPRQKQKTRHYFAAVERLWDYGMSSSPHVLNRAQSGSVQPKVQVQSEFTD 1740
 1760 GSTOPIYRGELEHGLGPGYIPAEVEDIMVTFRNOASRPYSFYSSLLSYSEDOQGA 1819
 1741 GSFTQPIYRGELEHGLGPGYIPAEVEDIMVTFRNOASRPYSFYSSLLSYSEDOQGA 1800
 1820 EPRKQFVKPNETKTYFWKVOHMAPTKDEPDCKAWAYFSDVLEKDVHSGLGIGLPLLYCHT 1879
 1801 EPRKQFVKPNETKTYFWKVOHMAPTKDEPDCKAWAYFSDVLEKDVHSGLGIGLPLLYCHT 1860
 1880 NTLNPAHGRQVTVQEFALFTTIPDETWSWPTENMERNCRAPCNQMEDPTFKENYFHA 1939
 1861 NTLNPAHGRQVTVQEFALFTTIPDETWSWPTENMERNCRAPCNQMEDPTFKENYFHA 1920
 1940 INGVIQMTLPLGLVNAQORIRWYLLSGSNENIHSIHPSGHVFTRKKEBYKMALYNLYP 1999
 1921 INGVIQMTLPLGLVNAQORIRWYLLSGSNENIHSIHPSGHVFTRKKEBYKMALYNLYP 1980
 2000 GVFTVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKCOTPLGMASGHIRDFQITAS 2059
 1981 GVFTVEMLPKAGIWRVECLIGELHAGMSTFLVYSNKCOTPLGMASGHIRDFQITAS 2040
 2060 GOYQOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTOGAROKFSSLYISQ 2119
 2041 GOYQOWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIHGIKTOGAROKFSSLYISQ 2100
 2120 FIIMYSLDGKWKQYRGNSCTGLMVFGNVDSSGIGKENIFNPPIIARVIRLHPHTHYSIRS 2179
 2101 FIIMYSLDGKWKQYRGNSCTGLMVFGNVDSSGIGKENIFNPPIIARVIRLHPHTHYSIRS 2160
 2180 TLRMELMGCDLNSCMLPLGMSKALSADAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2239
 2161 TLRMELMGCDLNSCMLPLGMSKALSADAQITASSYFTNMFTWSPSKARLHLQGRSNAWR 2220
 2240 PQVNNPKSWLQVDPQKTKWKTGVTITQGVKSLLTSMYVKEFLISSQDGHQWTLFPONGKV 2299
 2221 PQVNNPKSWLQVDPQKTKWKTGVTITQGVKSLLTSMYVKEFLISSQDGHQWTLFPONGKV 2280
 2300 KVFQGNQDSFTPVNSLDPPLILTRYLRTHYRTHQSWHQAIALRMEVLGCEAODLY 2351
 2281 KVFQGNQDSFTPVNSLDPPLILTRYLRTHYRTHQSWHQAIALRMEVLGCEAODLY 2332

RESULT 13

US-09-523-656-2
 ; Sequence 2, Application US/09523656
 ; Patent No. 6458563
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar S., John
 ; TITLE OF INVENTION: MODIFIED FACTOR VIII
 ; FILE REFERENCE: 75-951
 ; CURRENT APPLICATION NUMBER: US/09/523,656
 ; EARLIER APPLICATION DATE: 2000-03-10
 ; EARLIER APPLICATION NUMBER: 09/037,601
 ; EARLIER FILING DATE: 1998-03-10
 ; EARLIER APPLICATION NUMBER: 08/670,707
 ; EARLIER FILING DATE: 1996-06-26
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 2332
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-523-656-2

Query Match 98.9%; Score 12282; DB 4; Length 2332;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 20 ATRRYYLGAVELSWDYMOSDLGELPVDARPPRPVPSKPPNTSVVYKTLFVEFTDHLFN 79
 1 ATRRYYLGAVELSWDYMOSDLGELPVDARPPRPVPSKPPNTSVVYKTLFVEFTDHLFN 60

80 IAKPRPPNMGLLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYWKASGAEYDDOTSQ 139
 61 IAKPRPPNMGLLGPTTQAEVYDVTWITLKNMASHPVSLHAGVSVYWKASGAEYDDOTSQ 120
 140 REKEDDKVFPGGSHTYVWQVLKNGPMASDPLCLTYSYLSHVDLVKDLSGLIGALLVCR 199
 121 REKEDDKVFPGGSHTYVWQVLKNGPMASDPLCLTYSYLSHVDLVKDLSGLIGALLVCR 180
 200 EGSIAKEKTOITLHKFTILLPAVFPBGKSWHSETKNSLMQDRDAASARAWPKWHTVNGVNR 259
 181 EGSIAKEKTOITLHKFTILLPAVFPBGKSWHSETKNSLMQDRDAASARAWPKWHTVNGVNR 240
 260 SLPLGLICHERKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 319
 241 SLPLGLICHERKSVYVHVGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITLTAQTLL 300
 320 MDLGOFLLSCHISHQHDGMEAYVKVDSCEBEPQLIMKNEEAEYDDDLTDSMDVVR 379
 301 MDLGOFLLSCHISHQHDGMEAYVKVDSCEBEPQLIMKNEEAEYDDDLTDSMDVVR 360
 380 DDNSPSFIOIRSAVKKHPKTWVHYIAAEEDWDYAPLVLAPDDRYSYKSYLNNGPORIG 439
 361 DDNSPSFIOIRSAVKKHPKTWVHYIAAEEDWDYAPLVLAPDDRYSYKSYLNNGPORIG 420
 440 RYKVKVAFMAYTDTPTKTREAIQHESGILGPLLYGEVGDITLLIFKNOASRPYNIYPHGI 499
 421 RYKVKVAFMAYTDTPTKTREAIQHESGILGPLLYGEVGDITLLIFKNOASRPYNIYPHGI 480
 500 TDVAPLYSRRLPKGVKHLKDPFPLPGEIFKYKWTVTVEDEGFTKSDPCLTRYVYSSFYNME 559
 481 TDVAPLYSRRLPKGVKHLKDPFPLPGEIFKYKWTVTVEDEGFTKSDPCLTRYVYSSFYNME 540
 560 RDLASGLIGPLLI CYKESVQKGNQIMSDKRNVLFSVFDENRSWYLTENIQRELNPAG 619
 541 RDLASGLIGPLLI CYKESVQKGNQIMSDKRNVLFSVFDENRSWYLTENIQRELNPAG 600
 620 VQLEDPFQASNIMHSINGVYVDFSLQLSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 679
 601 VQLEDPFQASNIMHSINGVYVDFSLQLSVCLHEVAYWYIISIGAQTDFLSVFFSGYTFKH 660
 680 KQVYEDITLTFPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCDKMTGYE 739
 661 KQVYEDITLTFPFSGETVFMENPGLWILGCHNSDFRNGMTALLKVSSCDKMTGYE 720
 740 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTQKQFNATTIPENDIEKTDPFWAHRTMP 799
 721 DSYEDISAYLLSKNNAIEPRSFQNSRHPSTQKQFNATTIPENDIEKTDPFWAHRTMP 780
 800 KIQNVSSDLLMLLRQSPTPHGLSLSLQLOAKYETFSDDPSGAIIDNNSETHPRPQ 859
 781 KIQNVSSDLLMLLRQSPTPHGLSLSLQLOAKYETFSDDPSGAIIDNNSETHPRPQ 840
 860 LHSQSDMVFTEPSGLQRLNEKLTAAATELKLDFKVSSTSNLSTIPSDNLAAGTDN 919
 841 LHSQSDMVFTEPSGLQRLNEKLTAAATELKLDFKVSSTSNLSTIPSDNLAAGTDN 900
 920 TSSLGPPSPMPVHYDSQDITLFGKKSPLTBSGGPLSLSEENNDKSLLESGLMNSQSSSW 979
 901 TSSLGPPSPMPVHYDSQDITLFGKKSPLTBSGGPLSLSEENNDKSLLESGLMNSQSSSW 960
 980 GRNVSSSTESGRLFKGKRAHGPALLTKONALFKVSIKLTNKTNNNGATNKTHTIDGPSL 1039
 961 GRNVSSSTESGRLFKGKRAHGPALLTKONALFKVSIKLTNKTNNNGATNKTHTIDGPSL 1020
 1040 LIENSPSVWQNLISDTEFKKVTPLIHDRMLMDKNATALRLNHNKNTTSSKNMWTQOK 1099
 1021 LIENSPSVWQNLISDTEFKKVTPLIHDRMLMDKNATALRLNHNKNTTSSKNMWTQOK 1080
 1100 KEGTIPDPAQNPDMSEFFKMLFLPESARWIQTHGKNSLNSQGGSPKOLVSLGPEKSV 1159
 1081 KEGTIPDPAQNPDMSEFFKMLFLPESARWIQTHGKNSLNSQGGSPKOLVSLGPEKSV 1140
 1160 QNFLSEKKNVVGKGEFTKQVGLKEMVPPSSRNFLTNLMDLHNHNTHQEKKIQSEIEK 1219

Db 1141 QNFISEKKNVVGKGTVDGLKEMVFPSSRNLPJTNLNLHNNTHNOEKIQEIEIX 1200
QY 1220 KETLIQBNVVLPOHVTGKTNFKNKNIPLISTQNVEGSDGAYAPVLQFRLNDSTNR 1279
Db 1201 KETLIQBNVVLPOHVTGKTNFKNKNIPLISTQNVEGSDGAYAPVLQFRLNDSTNR 1260
QY 1280 TKKHTAHFSKKGREENLEGGNOTKQIVKRYACTTRISPNSTSOONFVTOQSKALKQFRL 1339
Db 1261 TKKHTAHFSKKGREENLEGGNOTKQIVKRYACTTRISPNSTSOONFVTOQSKALKQFRL 1320
QY 1340 PLEBTELEKRIIIVDDTSTOWSKNMKHLTPSLTQIDYNEKEKGAIITQSPISDCLTRSHSI 1399
Db 1321 PLEBTELEKRIIIVDDTSTOWSKNMKHLTPSLTQIDYNEKEKGAIITQSPISDCLTRSHSI 1380
QY 1400 POANRSLPFAKVSSPFSIRPIYILTRVLPQDNNSHLPAAASYRKDSQVQESSHPLOAKK 1459
Db 1381 POANRSLPFAKVSSPFSIRPIYILTRVLPQDNNSHLPAAASYRKDSQVQESSHPLOAKK 1440
QY 1460 NNLSLAILTLEMTGDQREVGLSGTSATNSVTYKXVENTVLPKPDLPKTSQKVELLPKVEI 1519
Db 1441 NNLSLAILTLEMTGDQREVGLSGTSATNSVTYKXVENTVLPKPDLPKTSQKVELLPKVEI 1500
QY 1520 YQKDLFPTETSNQSPGHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESSAKTPSK 1579
Db 1501 YQKDLFPTETSNQSPGHLDLVEGSLLOCTEGAIKXNEANRPGKVPFLRVATESSAKTPSK 1560
QY 1580 LLPLANDNHYGTQIPKEEWKQSKSEKPTAFKKKDTILSNACESHAAIINAGONKP 1639
Db 1561 LLPLANDNHYGTQIPKEEWKQSKSEKPTAFKKKDTILSNACESHAAIINAGONKP 1620
QY 1640 EIEVTWAKQRTERLCSQNPPLVLRHOREITRTTLQSDQEBIDYDDIISVEMKKEDFDIY 1699
Db 1621 EIEVTWAKQRTERLCSQNPPLVLRHOREITRTTLQSDQEBIDYDDIISVEMKKEDFDIY 1680
QY 1700 DEDNQSPRSQKTRRYFTAAVRELMDYGMSSSPHVLNRNQAQSGVPQPKVVFQFTD 1759
Db 1681 DEDNQSPRSQKTRRYFTAAVRELMDYGMSSSPHVLNRNQAQSGVPQPKVVFQFTD 1740
QY 1760 GSFTQPLRYGELNEHLGLGPIYRAEVEDNIMVTRNQAQSPFPYSSLSIYSESDQOQA 1819
Db 1741 GSFTQPLRYGELNEHLGLGPIYRAEVEDNIMVTRNQAQSPFPYSSLSIYSESDQOQA 1800
QY 1820 EPRKNFVKNETKYFWKVQHMAPTKDEPCKAWAYFSDVLEKDVHSGLGIGLLVCHT 1879
Db 1801 EPRKNFVKNETKYFWKVQHMAPTKDEPCKAWAYFSDVLEKDVHSGLGIGLLVCHT 1860
QY 1880 NTLNPAHGRQVTVORPALFFTTIPDETTSWYFTENMERNCRAPCNIQMEDDPTFKENYRFHA 1939
Db 1861 NTLNPAHGRQVTVORPALFFTTIPDETTSWYFTENMERNCRAPCNIQMEDDPTFKENYRFHA 1920
QY 1940 INGVIMDTPLGLVNAQDQRIRWYLLSGNSNENIHSIHFSGHVTVVRKXEYKMALYNLYP 1999
Db 1921 INGVIMDTPLGLVNAQDQRIRWYLLSGNSNENIHSIHFSGHVTVVRKXEYKMALYNLYP 1980
QY 2000 GVFTVEMLPKAGIWRVECHLGHAGMETLFLVYSNKCQTPLGMASGHIRDFQITAS 2059
Db 1981 GVFTVEMLPKAGIWRVECHLGHAGMETLFLVYSNKCQTPLGMASGHIRDFQITAS 2040
QY 2060 GQYQWAPKLABLHYSGSINAWSKPEPSWIKVDLLAPMIHGIKTQAROKFSSLIYSQ 2119
Db 2041 GQYQWAPKLABLHYSGSINAWSKPEPSWIKVDLLAPMIHGIKTQAROKFSSLIYSQ 2100
QY 2120 FIIMYSLOGKQWOTYRGNSTGFLMYFFGVDSGGIKHNIENPPIIARYLRHPTHYSIRS 2179
Db 2101 FIIMYSLOGKQWOTYRGNSTGFLMYFFGVDSGGIKHNIENPPIIARYLRHPTHYSIRS 2160
QY 2180 TLRBELMGCDLNSCSMPLGMSKAIISDAQITASSYFTNNFATWSPSKARLHQGSNAWR 2239
Db 2161 TLRBELMGCDLNSCSMPLGMSKAIISDAQITASSYFTNNFATWSPSKARLHQGSNAWR 2220
QY 2240 PQVNNPKWLQVDFQKTMKVTVGTQGVKSLTSTMYVKEFLISSQDQHQWTLFFQNGKV 2299

Db 2221 PQVNNPKWLQVDFQKTMKVTVGTQGVKSLTSTMYVKEFLISSQDQHQWTLFFQNGKV 2280
QY 2300 KVFQGNQDSPTPVNSLDPPLLTRYLRIHQSWVHQAIALAMEVLGCEAQLY 2351
Db 2281 KVFQGNQDSPTPVNSLDPPLLTRYLRIHQSWVHQAIALAMEVLGCEAQLY 2332
RESULT 14
PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Ruge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03275
; FILING DATE: 19930407
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: BMU 106PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cdna sequence
PCT-US93-03275-4
Query Match 98.9%; Score 12282; DB 5; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 20 ATRYIYGAVELSDVNDYQSDIGELPVDARPPRPVKSPFNWTSVYKTLFVETDHLFN 79
Db 1 ATRYIYGAVELSDVNDYQSDIGELPVDARPPRPVKSPFNWTSVYKTLFVETDHLFN 60
QY 80 IAKRPPMGLGPTIOAEVYDVTVTILKNASHPVSLHAGVSYWKASGAEYDDQTSQ 139
Db 61 IAKRPPMGLGPTIOAEVYDVTVTILKNASHPVSLHAGVSYWKASGAEYDDQTSQ 120
QY 140 REKEDDKVFGGSHYVQVQLKENGPMASPLCLITYSLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFGGSHYVQVQLKENGPMASPLCLITYSLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKKTQTLHKFILLFAVFDGKSHSETKNSLMQDRDAASARAWPKMHTVNGYNR 259

181 EGSFAKENTQTLHKFILLFANVDECKSWHSETKNSLQORDAASARAPKWHVNGYNR 240
260 SLPLGILGCHRSYVWHVITGMETTPPVHISIFLEGHTFLVNRHQASLEISPTIFFTAQTL 319
241 SLPLGILGCHRSYVWHVITGMETTPPVHISIFLEGHTFLVNRHQASLEISPTIFFTAQTL 300
320 MDLQGLLSCCHSHCHDGMAYVVDSCPEPEQLIMKNEAEADYDDDLTDSMDVVR 379
301 MDLQGLLFCCHSHCHDGMAYVVDSCPEPEQLIMKNEAEADYDDDLTDSMDVVR 360
380 DDNSPSPQIRSVAKKHPKVVHYIAAEEDWDYAPLVLPADDDRSYKSOYLNNGPQRIG 439
361 DDNSPSPQIRSVAKKHPKVVHYIAAEEDWDYAPLVLPADDDRSYKSOYLNNGPQRIG 420
440 RYKVKVRNAYTDETFKTRREALQHESGILGPLLICEVGETLLIIFPKQAASRPYNIYPHGI 499
421 RYKVKVRNAYTDETFKTRREALQHESGILGPLLICEVGETLLIIFPKQAASRPYNIYPHGI 480
500 TVRPLYSRRLPKGVKHLKDPPLPGETPFKYKWTVTVEDGPTKSDPRCLTRYSSFVWME 559
481 TVRPLYSRRLPKGVKHLKDPPLPGETPFKYKWTVTVEDGPTKSDPRCLTRYSSFVWME 540
560 RDLASGLIGPLLICYESVDQGNQIMSDKRNVLFSVFCNRSWYLTENIQREFLPNPA 619
541 RDLASGLIGPLLICYESVDQGNQIMSDKRNVLFSVFCNRSWYLTENIQREFLPNPA 600
620 VOLEDPRQASNIHISINGYVFDLSIOLSVCLHEVAYWVILSIGAQDTFLSVFFSGYTFFKH 679
601 VOLEDPRQASNIHISINGYVFDLSIOLSVCLHEVAYWVILSIGAQDTFLSVFFSGYTFFKH 660
680 KMVYEDTLTPFPFSGETVPMENPGLMILGCHNSDFNRNGWTALLKVSSODKNTGDYYE 739
661 KMVYEDTLTPFPFSGETVPMENPGLMILGCHNSDFNRNGWTALLKVSSODKNTGDYYE 720
740 DSYEDISAYLSKNAIPEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFWAHRTMP 799
721 DSYEDISAYLSKNAIPEPRFSQNSRHPSTRQKQFNATTIPENDIEKTDPFWAHRTMP 780
800 KIONVSSDLMLLRQSTPHGLSLSDIOEAKYETFSDDPGALDSNNSISEMTHPRPQ 859
781 KIONVSSDLMLLRQSTPHGLSLSDIOEAKYETFSDDPGALDSNNSISEMTHPRPQ 840
860 LHSQDMVFTPESGIQLRINKLGTATATELKKLDFKVSSTSNLIISTIPDNLAAGTDN 919
841 LHSQDMVFTPESGIQLRINKLGTATATELKKLDFKVSSTSNLIISTIPDNLAAGTDN 900
920 TSSIGPSPMPVHYDSQDITTLFGKKSPLITRESGGPLSISEENNDKLLSGLMNSQESSW 979
901 TSSIGPSPMPVHYDSQDITTLFGKKSPLITRESGGPLSISEENNDKLLSGLMNSQESSW 960
980 GKNVSSSTESGRLFKGRAGPALLTKDNALFKVSI SLLKTKNTKSNNSATNRKTHIDGPSL 1039
961 GKNVSSSTESGRLFKGRAGPALLTKDNALFKVSI SLLKTKNTKSNNSATNRKTHIDGPSL 1020
1040 LIENSPPYQWNILEDSTEFKVTPLIHDRMLMDKNATALRLNHNKNTTSKKNMEMWQOK 1099
1021 LIENSPPYQWNILEDSTEFKVTPLIHDRMLMDKNATALRLNHNKNTTSKKNMEMWQOK 1080
1100 KEGPIPPDAQNDPMGFFKMLFLPESARWIORTHGKNSLNSGGPSPKQIVSLGPSPKSV 1159
1081 KEGPIPPDAQNDPMGFFKMLFLPESARWIORTHGKNSLNSGGPSPKQIVSLGPSPKSV 1140
1160 QNFLEKXKVVVGKEFTKDVGLKEMVFPSSNLFLTLNDLNHNHNTHNCKKIOEEIEK 1219
1141 QNFLEKXKVVVGKEFTKDVGLKEMVFPSSNLFLTLNDLNHNHNTHNCKKIOEEIEK 1200
1220 KETLIQENVLPQIHTVGTGKNFMKNLFLLSRQNVESYDGAAPVQDPRSLNDSNR 1279
1201 KETLIQENVLPQIHTVGTGKNFMKNLFLLSRQNVESYDGAAPVQDPRSLNDSNR 1260
1280 TKKHTAHSKKEEENLEGLGNQTKQIVKYACTTRISPNTSQOQNFVTRQSRKALKQFRL 1339

RESULT 15
PCT-US94-13200-2

1261 TKKHTAHSKKEEENLEGLGNQTKQIVKYACTTRISPNTSQOQNFVTRQSRKALKQFRL 1320
1340 PLEETBELKRIIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKEKGALTOQSPISDCLTRHSHI 1399
1321 PLEETBELKRIIIVDDTSTOWSKNMKHLTPSTLTQIDYNEKEKGALTOQSPISDCLTRHSHI 1380
1400 PQANRSLPLAKVSSFPISIRPIYLVTRVLFQDNSHLPAAASYRKXKDSGVQZESSHFLQAGAK 1459
1381 PQANRSLPLAKVSSFPISIRPIYLVTRVLFQDNSHLPAAASYRKXKDSGVQZESSHFLQAGAK 1440
1460 NNLSLAILTLEMTQDQREVSGSLGTSATNSVYTKKVENTVLPKPLPKTSGKVELLPKVHI 1519
1441 NNLSLAILTLEMTQDQREVSGSLGTSATNSVYTKKVENTVLPKPLPKTSGKVELLPKVHI 1500
1520 YQXDLPTETSNQSPGHLDLVEGSLLOGTGAHKWNEANRPGKVPFLRVATESSAKTPSK 1579
1501 YQXDLPTETSNQSPGHLDLVEGSLLOGTGAHKWNEANRPGKVPFLRVATESSAKTPSK 1560
1580 LLDPLANDNHYGTQIPKEEWKSOEKSPKTAFAKKDKTILSLNACESNHAAIAINEGQNK 1639
1561 LLDPLANDNHYGTQIPKEEWKSOEKSPKTAFAKKDKTILSLNACESNHAAIAINEGQNK 1620
1640 EIEVTWAKQGTRELCSQNPVLPKXHORBITRTTLOSQDBEIDYDDTISIVEMKKEDEIY 1699
1621 EIEVTWAKQGTRELCSQNPVLPKXHORBITRTTLOSQDBEIDYDDTISIVEMKKEDEIY 1680
1700 DEENQSPRSFQKTRHYFTAAVERLWDYGNSSSPHVRNEAQSGSVPOKVVVQFETD 1759
1681 DEENQSPRSFQKTRHYFTAAVERLWDYGNSSSPHVRNEAQSGSVPOKVVVQFETD 1740
1760 GSFTQPLVYRGELNHLGLLGPYIRAEVEDNIMVTRNQASRPYFYSLLISYBBDQOQGA 1819
1741 GSFTQPLVYRGELNHLGLLGPYIRAEVEDNIMVTRNQASRPYFYSLLISYBBDQOQGA 1800
1820 EPRNFXVKNETKTYFKVQVHHMPTKDEPDCKAWAYPSDVLBKVHSGILGILLVCHT 1879
1801 EPRNFXVKNETKTYFKVQVHHMPTKDEPDCKAWAYPSDVLBKVHSGILGILLVCHT 1860
1880 NTLNPAHGRQVTVQEFALFTTIDETKSWYPTENNERNCRAPCNIOMEDPTFKENYRPHA 1939
1861 NTLNPAHGRQVTVQEFALFTTIDETKSWYPTENNERNCRAPCNIOMEDPTFKENYRPHA 1920
1940 INGVIMDTPLGLVMAQDQIRWYLLSNGSNIHISIFSGHVFTVRKKEEYKMALYNLYP 1999
1921 INGVIMDTPLGLVMAQDQIRWYLLSNGSNIHISIFSGHVFTVRKKEEYKMALYNLYP 1980
2000 GVPETVEMLPKAGIWRVECLIGEHLFAGMSTLPLVYSNKCOTPLGKASGHIRDFQITAS 2059
1981 GVPETVEMLPKAGIWRVECLIGEHLFAGMSTLPLVYSNKCOTPLGKASGHIRDFQITAS 2040
2060 GOYGOWAPKLARLHYSGSINAWSTKPPSWIKVDLLAPMIHGIKTQAROKPSLSYISQ 2119
2041 GOYGOWAPKLARLHYSGSINAWSTKPPSWIKVDLLAPMIHGIKTQAROKPSLSYISQ 2100
2120 FIIMYSLDGKXQWYRGNSTGTLVFFGNVDSSGKINENPPIIARYIRLHPHTYSIRS 2179
2101 FIIMYSLDGKXQWYRGNSTGTLVFFGNVDSSGKINENPPIIARYIRLHPHTYSIRS 2160
2180 TLRVLMGCDLNSCSMPLGMESKAISDAQTASSYFTNNFATWSPSKARLHLQGRSNAWR 2239
2161 TLRVLMGCDLNSCSMPLGMESKAISDAQTASSYFTNNFATWSPSKARLHLQGRSNAWR 2220
2240 PQVNNPEWLOVDQKMKVGTVTQGVKSLLTSMYVKEFLISSSQDGHOWTLFPQNGKV 2299
2221 PQVNNPEWLOVDQKMKVGTVTQGVKSLLTSMYVKEFLISSSQDGHOWTLFPQNGKV 2280
2300 KVFQGNQDSFPPVNSLDPPLLTPLYRILHPQSWVHQIALRMEVLGCEAQDLY 2351
2281 KVFQGNQDSFPPVNSLDPPLLTPLYRILHPQSWVHQIALRMEVLGCEAQDLY 2332

```

; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
; PCT-US94-13200-2

Query Match 98.9%; Score 12282; DB 5; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 20 ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTDHLFN 79
Db 1 ATRRYILGAVELSDYMQSDLGELPVDARPPRPVKSPFPNTSVVYKTLFVEFTDHLFN 60

Qy 80 IAKPRPMMGLGPTIOAEVYDVTVITLKNASHFVSLHAGVSVYKASGAEAYDDQTSQ 139
Db 61 IAKPRPMMGLGPTIOAEVYDVTVITLKNASHFVSLHAGVSVYKASGAEAYDDQTSQ 120

Qy 140 REKDDKVFPGSGSTYYQVQLKENGPMASDPLCLITYSLSHVDVLKQDINSGLICALLVCR 199
Db 121 REKDDKVFPGSGSTYYQVQLKENGPMASDPLCLITYSLSHVDVLKQDINSGLICALLVCR 180

Qy 200 EGSIAKERTQTLHKPILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKGHTYGVNR 259
Db 181 EGSIAKERTQTLHKPILLFAVFDGKSWHSTKNSLMQDRDAASARAWPKGHTYGVNR 240

Qy 260 SLPLGIGCHRSVYWHVIGMGTTEVEHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 319
Db 241 SLPLGIGCHRSVYWHVIGMGTTEVEHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTL 300

Qy 320 MDLQCFLLSCHSHQHDGMEAYVKVDCSCEPQOLIMQNEEASDYDDLTDSDMDVVRP 379
Db 301 MDLQCFLLSCHSHQHDGMEAYVKVDCSCEPQOLIMQNEEASDYDDLTDSDMDVVRP 360

Qy 380 DDDNSPSFIQIRSVAKGPKTWHTVHYIAAEEDWDYAPLVLPDDRYSKYSOYLNNGPQIG 439

```

QY 1520 YQDLFPPTETSGSPGHLDLVEGSLLOTEGAIKWNEANRCKVPFLRVATESSAKTPSK 1579
DB 1501 YQDLFPPTETSGSPGHLDLVEGSLLOTEGAIKWNEANRCKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLANDMNYGTQIPKEWKSQBSPEKTAFFKKTILSNACESHAAIAINEGQNK 1639
DB 1561 LLDPLANDMNYGTQIPKEWKSQBSPEKTAFFKKTILSNACESHAAIAINEGQNK 1620
QY 1640 BIEVWAKQGTERTLCSQNPVLRKHOREITRTTLOSQOEIDYDDTISVEMKXEDPDIY 1699
DB 1621 BIEVWAKQGTERTLCSQNPVLRKHOREITRTTLOSQOEIDYDDTISVEMKXEDPDIY 1680
QY 1700 DEDENQSPESFQKTRHFIAAVERLWDYGNSSPHVLNRNAQSGSVPOFKKVVFOEFTD 1759
DB 1681 DEDENQSPESFQKTRHFIAAVERLWDYGNSSPHVLNRNAQSGSVPOFKKVVFOEFTD 1740
QY 1760 GSFTQPLYRGELNEHGLGPGYIRAEVEDNIMVTPRNOASRPYSPYSSLSIYEBDQROGA 1819
DB 1741 GSFTQPLYRGELNEHGLGPGYIRAEVEDNIMVTPRNOASRPYSPYSSLSIYEBDQROGA 1800
QY 1820 EPRKNFVKPNETKTYFWKVOHHMPTKDEFCCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
DB 1801 EPRKNFVKPNETKTYFWKVOHHMPTKDEFCCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHA 1939
DB 1861 NTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRPHA 1920
QY 1940 INGYIMDTLPGLVMAQDORIRWYLLSMGSENENIHSIHFSCHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGYIMDTLPGLVMAQDORIRWYLLSMGSENENIHSIHFSCHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVFEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYNSKCOOTPLGMASGHIRDFOITAS 2059
DB 1981 GVFEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYNSKCOOTPLGMASGHIRDFOITAS 2040
QY 2060 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKFSLSYISQ 2119
DB 2041 GOYGOWAPKLARLHYSGSINAWSTKEPFSWKVDLLAPMIHGIKTQGARQKFSLSYISQ 2100
QY 2120 FIIMYSLDGKKWQTYRGNSTGTLNMFVFGNVDSSGIXKHNIENPPIIARYIRLHPTYSIRS 2179
DB 2101 FIIMYSLDGKKWQTYRGNSTGTLNMFVFGNVDSSGIXKHNIENPPIIARYIRLHPTYSIRS 2160
QY 2180 TURBELMGCDLNSCMLGMBKASDAQITASSYFTNMFAWSPSKARLHLQGSNAWR 2239
DB 2161 TURBELMGCDLNSCMLGMBKASDAQITASSYFTNMFAWSPSKARLHLQGSNAWR 2220
QY 2240 POVNNPKWLQVDFQKTMKVTGVTQGVKSLLTSMYKVEFLISSQDGHQWTLFPQNGKV 2299
DB 2221 POVNNPKWLQVDFQKTMKVTGVTQGVKSLLTSMYKVEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KYFOGNQDSFTPVNSLDPPLILTRVLRHPOSWVHQIALRMEVLGCEAODLY 2351
DB 2281 KYFOGNQDSFTPVNSLDPPLILTRVLRHPOSWVHQIALRMEVLGCEAODLY 2332

Search completed: April 13, 2004, 14:08:32
Job time : 31.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2004, 14:08:32 ; Search time 61.5 Seconds
(without alignments)

10051.498 Million cell updates/sec

Title: NP000123-328-355-581

Perfect score: 12415

Sequence: 1 MQIELSTCFCLLRFCFSA.....VWQIALRMEVLGCEAQDLY 2351

Scoring table: 3LOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues

Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US6C_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US6D_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12399	99.9	2351	14	US-10-132-829-4
2	12399	99.9	2351	14	Sequence 4, Appli
3	12399	99.9	2351	14	Sequence 27, Appl
4	12399	99.8	2351	12	US-10-133-907-4
5	12399	99.8	2351	12	US-10-411-037-30
6	12399	99.8	2351	12	US-10-411-026-30
7	12399	99.8	2351	15	US-10-360-101-229
8	12399	99.9	2332	14	US-09-957-641-2
9	12399	99.9	2332	14	Sequence 2, Appli
10	12399	99.9	2332	14	US-10-131-510A-2
11	12399	99.9	2332	14	US-10-445-235-2
12	12399	99.9	2332	16	US-10-239-498A-2
13	12399	99.9	2332	16	US-10-187-319-6
14	12399	99.9	2332	14	US-10-187-319-6
15	12399	99.9	2332	14	US-10-187-319-6

16	7393	59.5	1471	12	US-10-681-970-2	Sequence 2, Appli
17	7393	59.5	1471	13	US-10-095-718-2	Sequence 2, Appli
18	7294	58.6	1459	16	US-10-239-498A-13	Sequence 13, Appli
19	7284	58.7	1459	16	US-10-239-498A-15	Sequence 15, Appli
20	7281	58.6	1459	16	US-10-239-498A-4	Sequence 4, Appli
21	7208	58.1	1438	13	US-10-006-091-1	Sequence 1, Appli
22	7208	58.1	1438	13	US-10-047-257-1	Sequence 1, Appli
23	7208	58.1	1438	14	US-10-225-900-1	Sequence 1, Appli
24	6253	50.4	1431	12	US-10-681-970-4	Sequence 4, Appli
25	6253	50.4	1431	13	US-10-095-718-4	Sequence 4, Appli
26	6103	49.2	1443	14	US-10-187-319-39	Sequence 39, Appli
27	6103	49.2	1443	14	US-10-131-510A-39	Sequence 39, Appli
28	2825.5	22.7	2224	14	US-10-115-563-14	Sequence 14, Appli
29	2823.5	22.7	2196	15	US-10-360-101-259	Sequence 259, App
30	2823.5	22.7	2224	14	US-10-172-712-31	Sequence 31, Appli
31	1723	13.9	368	14	US-10-187-319-4	Sequence 4, Appli
32	1723	13.9	368	14	US-10-131-510A-4	Sequence 4, Appli
33	1340.5	10.8	1160	12	US-10-147-493-234	Sequence 234, App
34	1340.5	10.8	1160	12	US-10-145-127-234	Sequence 234, App
35	1340.5	10.8	1160	12	US-10-160-503-234	Sequence 234, App
36	1340.5	10.8	1160	12	US-10-143-118-234	Sequence 234, App
37	1340.5	10.8	1160	12	US-10-144-993-234	Sequence 234, App
38	1340.5	10.8	1160	12	US-10-158-787-234	Sequence 234, App
39	1340.5	10.8	1160	12	US-10-140-024-234	Sequence 234, App
40	1340.5	10.8	1160	14	US-10-028-072-234	Sequence 234, App
41	1340.5	10.8	1160	14	US-10-121-049-234	Sequence 234, App
42	1340.5	10.8	1160	14	US-10-123-904-234	Sequence 234, App
43	1340.5	10.8	1160	14	US-10-140-470-234	Sequence 234, App
44	1340.5	10.8	1160	14	US-10-175-748-234	Sequence 234, App
45	1340.5	10.8	1160	14	US-10-176-918-234	Sequence 234, App

ALIGNMENTS

RESULT 1
US-10-132-829-4
; Sequence 4, Application US/10132829
; Publication No. US20030044982A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: with vesicle vector
; FILE REFERENCE: 6627-PAL170
; CURRENT APPLICATION NUMBER: US/10/132,829
; CURRENT FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-132-829-4

Query Match		99.9%	Score 12399;	DB 14;	Length 2351;
Best Local Similarity		99.9%	Pred. No. 0;		
Matches 2348;		Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;
Qy	1	MQIELSTCFCLLRFCFSATRRYYLGAVELSDYMSQDLGELPVDARPPRPVKSPFN	60		
Db	1	MQIELSTCFCLLRFCFSATRRYYLGAVELSDYMSQDLGELPVDARPPRPVKSPFN	60		
Qy	61	TSVYKKTLEVFETHLPNIAKPRPPMGLLGTIQAEVDTVVITLKNASHPVLSHAV	120		
Db	61	TSVYKKTLEVFETHLPNIAKPRPPMGLLGTIQAEVDTVVITLKNASHPVLSHAV	120		
Qy	121	GVSYWKASGAEYDDQTSOREKDDKVPFGSHYVWVLKENGPMASDPLCLTYSLSH	180		
Db	121	GVSYWKASGAEYDDQTSOREKDDKVPFGSHYVWVLKENGPMASDPLCLTYSLSH	180		

181 VDLVKDNLNSGLIGALLVCREGSLAKEXTQTLHKFILLFAVDEGSKWSESTKNSLMQDRD 240
181 VDLVKDNLNSGLIGALLVCREGSLAKEXTQTLHKFILLFAVDEGSKWSESTKNSLMQDRD 240
241 AASARAWPKXHTVNGYVYNNRSLPGLIGCHRSVYVHVHVGMTTPVHGFPLEGHTFLVRNH 300
241 AASARAWPKXHTVNGYVYNNRSLPGLIGCHRSVYVHVHVGMTTPVHGFPLEGHTFLVRNH 300
301 QOASLEISPIELTAOTLMLDGLGOFILCHLSSHOHDGMEAYVAVDSCPPEPOLIMKXNE 360
301 QOASLEISPIELTAOTLMLDGLGOFILCHLSSHOHDGMEAYVAVDSCPPEPOLIMKXNE 360
361 EAEDYDDDLTDSMDVVRFDNDNSPFTQIRSAVAKHFKTWVHYIAAEEDWDYAPLVLA 420
361 EAEDYDDDLTDSMDVVRFDNDNSPFTQIRSAVAKHFKTWVHYIAAEEDWDYAPLVLA 420
421 PDRSYKQVILNNGPQIRGRYKVKRMAVYDETFKREAIQHSIGLGPLLYGEVGDTL 480
421 PDRSYKQVILNNGPQIRGRYKVKRMAVYDETFKREAIQHSIGLGPLLYGEVGDTL 480
481 LIIFKNOASREYNILYPHGITDVRPLYSRRLPKGVKHLKDFILPGEIFPKYKWTVTVEDGP 540
481 LIIFKNOASREYNILYPHGITDVRPLYSRRLPKGVKHLKDFILPGEIFPKYKWTVTVEDGP 540
541 TKSDPRCLTRYSSFVAMERDLASGLIGPLLYCYKESVDQKNOIVSDKRNVLIFSVPDE 600
541 TKSDPRCLTRYSSFVAMERDLASGLIGPLLYCYKESVDQKNOIVSDKRNVLIFSVPDE 600
601 NRSYLTENIORFLPNPAGVQLEBPPEFOASNIMESINGVYFDSLSVCLHEVAYWYLS 660
601 NRSYLTENIORFLPNPAGVQLEBPPEFOASNIMESINGVYFDSLSVCLHEVAYWYLS 660
661 ICAOTDFLSVPFSGVTFKHKVYEDTLTLFPFSGBTVMSENGLITLGHNSDFNRG 720
661 ICAOTDFLSVPFSGVTFKHKVYEDTLTLFPFSGBTVMSENGLITLGHNSDFNRG 720
721 MTALLKVSSCDKGTDDYEDSYEDISAYLLSKNAIEPRFSQNSRHPSTKQFNATTI 780
721 MTALLKVSSCDKGTDDYEDSYEDISAYLLSKNAIEPRFSQNSRHPSTKQFNATTI 780
781 PENDIEKTDPAFHRTMPKTIQNTSSDILLMLLQSPPHGLISLSDIOBAKYETFSDDPS 840
781 PENDIEKTDPAFHRTMPKTIQNTSSDILLMLLQSPPHGLISLSDIOBAKYETFSDDPS 840
841 PGADSNNLSGEMTHFRQLHSGDMVTPPSGLQLRLNEXLGTAAATELKKLPFKVSST 900
841 PGADSNNLSGEMTHFRQLHSGDMVTPPSGLQLRLNEXLGTAAATELKKLPFKVSST 900
901 SNNLISTIPSDNLAAGTONTSLGPPSPVHYDSQDITTLFGKSSPLTESGGPLSLSEE 960
901 SNNLISTIPSDNLAAGTONTSLGPPSPVHYDSQDITTLFGKSSPLTESGGPLSLSEE 960
961 NNDKSLLESGLMNSOESSWGKNVSTESGRLPKGRAGHPALLTKDNALFKVYSLLKTN 1020
961 NNDKSLLESGLMNSOESSWGKNVSTESGRLPKGRAGHPALLTKDNALFKVYSLLKTN 1020
1021 KTSNNSATNRKTHIDCPILLSIENSFVQNILESPTFKVVTPLIHDRMLMDKNAATALRL 1080
1021 KTSNNSATNRKTHIDCPILLSIENSFVQNILESPTFKVVTPLIHDRMLMDKNAATALRL 1080
1081 NMSNKTTSKKNMVMVQKKEGPIPPDAQNPMSFPKMLFIPESARWQRTHGKNSLNSG 1140
1081 NMSNKTTSKKNMVMVQKKEGPIPPDAQNPMSFPKMLFIPESARWQRTHGKNSLNSG 1140
1141 QGPPSPKQLVSLGPEKSVQGNFLSBKNKVUVKGEBFTKDVGLKEMVFPSSRNLFITNLDN 1200
1141 QGPPSPKQLVSLGPEKSVQGNFLSBKNKVUVKGEBFTKDVGLKEMVFPSSRNLFITNLDN 1200
1201 LHEANTHNOEKKIQBSIEKKTLLIQENVVLPIQHTVTGTFKNFMKNLFLLLSTRQNVESYD 1260
1201 LHEANTHNOEKKIQBSIEKKTLLIQENVVLPIQHTVTGTFKNFMKNLFLLLSTRQNVESYD 1260
1261 GAYAPVQLDPRSLNDSNTKTKHTAHFSKKGEBEENLEGLNQTKQIVEKYACTTRISPN 1320

1261 GAYAPVQLDPRSLNDSNTKTKHTAHFSKKGEBEENLEGLNQTKQIVEKYACTTRISPN 1320
1321 SQQNFTQBSKRALQKQRIPLPLETELEKRIIVDDTSTQNSKNMKHLTPSTLTQIDVNEKE 1380
1321 SQQNFTQBSKRALQKQRIPLPLETELEKRIIVDDTSTQNSKNMKHLTPSTLTQIDVNEKE 1380
1381 KGAITOSPLSDCLTRSHSIPQANRSPLPIAKVSPFSIRPIYLRVLFQDNSSHLPAASY 1440
1381 KGAITOSPLSDCLTRSHSIPQANRSPLPIAKVSPFSIRPIYLRVLFQDNSSHLPAASY 1440
1441 RKXDSGVQESSHFLQAKAKNLSIALITLMTQDQREVGSLGTSATNSVYTKVENVTLVP 1500
1441 RKXDSGVQESSHFLQAKAKNLSIALITLMTQDQREVGSLGTSATNSVYTKVENVTLVP 1500
1501 KPDLPKTSGVKEVLLPKVHIYQKDLFPETNSGSPGHLDLVEGSLLOCTEGAIKWNENRP 1560
1501 KPDLPKTSGVKEVLLPKVHIYQKDLFPETNSGSPGHLDLVEGSLLOCTEGAIKWNENRP 1560
1561 GKVPFLRVATESSAKTPSKLLDPLANDNHYGTQIPKEEWSQEKSPKTAFFKXDTILSL 1620
1561 GKVPFLRVATESSAKTPSKLLDPLANDNHYGTQIPKEEWSQEKSPKTAFFKXDTILSL 1620
1621 NACESNHAIIAINEGQNKPEIEVTWAKQGTRELCSONPPVLRKHQREITRTTLOSQEE 1680
1621 NACESNHAIIAINEGQNKPEIEVTWAKQGTRELCSONPPVLRKHQREITRTTLOSQEE 1680
1681 IDYDDTISVEMKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
1681 IDYDDTISVEMKEDFDIYDEDENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHVLNR 1740
1741 AQSGSVPOFKKVPVQFTDGSFTQPLYRGELNHLGLLGYIRAEVEDNINMTFRNOASR 1800
1741 AQSGSVPOFKKVPVQFTDGSFTQPLYRGELNHLGLLGYIRAEVEDNINMTFRNOASR 1800
1801 PYSYSSLIISYEBDOROGASPRKNFKNETKIYFWKVQHMAPTKDEFFCKAWAYSDV 1860
1801 PYSYSSLIISYEBDOROGASPRKNFKNETKIYFWKVQHMAPTKDEFFCKAWAYSDV 1860
1861 DLEKXDSHGLIGPLLVTCHTNTLPAHQRTVQVDFALFFTFIDETKSWYFTENNERCRA 1920
1861 DLEKXDSHGLIGPLLVTCHTNTLPAHQRTVQVDFALFFTFIDETKSWYFTENNERCRA 1920
1921 PCNTQMBDPTFKENYRPFALNGYIMDTPLGLVMAQDQRIRWYLLSMGSENENIHSIFSGH 1980
1921 PCNTQMBDPTFKENYRPFALNGYIMDTPLGLVMAQDQRIRWYLLSMGSENENIHSIFSGH 1980
1981 VFTVRKKEEKWALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYGNKC 2040
1981 VFTVRKKEEKWALYNLYPGVFETVEMLPKAGIWRVECLIGELHAGMSTFLVYGNKC 2040
2041 QTPLGASGHTRDFQITASGOYQWAPKPLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2041 QTPLGASGHTRDFQITASGOYQWAPKPLARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
2101 HGIKTQARQXFSLLYSQFIIMYSLDGKKQVTRGNSGTGLMVPFGNVDSGSKHNIEN 2160
2101 HGIKTQARQXFSLLYSQFIIMYSLDGKKQVTRGNSGTGLMVPFGNVDSGSKHNIEN 2160
2161 PPIIARYRLPHPTHYSIRSTRLXELMGCNLSNPLGMESKAISSDAQITASSYFTNNFA 2220
2161 PPIIARYRLPHPTHYSIRSTRLXELMGCNLSNPLGMESKAISSDAQITASSYFTNNFA 2220
2221 TWSFSKARLHLQGRSNARPOVNNPKEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
2221 TWSFSKARLHLQGRSNARPOVNNPKEWLQVDFQKTMKVTVGTTQGVKSLTSMYVKEFL 2280
2281 ISSSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRHPQSWHQAIALRM 2340
2281 ISSSQDGHQWTLFTQNGKVKVQGNQDSFTPVNSLDPPLLTRYLRHPQSWHQAIALRM 2340
2341 EVLGCERQDLY 2351

Db 2341 EVLGCRAQDLY 2351

RESULT 2

US-10-172-712-27

; Sequence 27, Application US/10172712

; Publication No. US20030125232A1

; GENERAL INFORMATION:

; APPLICANT: GRIFFIN, JOHN H.

; APPLICANT: GALE, ANDREW J.

; APPLICANT: GETZOFF, ELIZABETH D.

; APPLICANT: PELLEQUER, JEAN-LUC

; TITLE OF INVENTION: STABILIZED PROTEINS WITH ENGINEERED DISULFIDE BONDS

; FILE REFERENCE: 4198-4001U51

; CURRENT APPLICATION NUMBER: US/10/172,712

; CURRENT FILING DATE: 2002-09-30

; PRIOR APPLICATION NUMBER: 60/298,578

; PRIOR FILING DATE: 2001-06-14

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 27

; LENGTH: 2351

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-172-712-27

Query Match 99.9%; Score 12399; DB 14; Length 2351;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MQIELSTCFELCLLRFCSATRRYYLGAVELSWDYMOSDLGELVDARFPFPRVPKSPFPN 60

Db 1 MQIELSTCFELCLLRFCSATRRYYLGAVELSWDYMOSDLGELVDARFPFPRVPKSPFPN 60

Qy 61 TSVVYKTKLTFVEFTHLFLNIAKRPMPMGLLGPTIQAEVYTVITLKNMASHPVSLHAV 120

Db 61 TSVVYKTKLTFVEFTHLFLNIAKRPMPMGLLGPTIQAEVYTVITLKNMASHPVSLHAV 120

Qy 121 GVSVMKASEGAEYDDQTSOREKDDKVPFGGSHYVMVOLKENGPMASDPLCLTYSLH 180

Db 121 GVSVMKASEGAEYDDQTSOREKDDKVPFGGSHYVMVOLKENGPMASDPLCLTYSLH 180

Qy 181 VDLVKDNLNSGLIGALLVCREGLAKEKQTQLHKEITLFAVPEDEKSMHSETKNSLMQDRD 240

Db 181 VDLVKDNLNSGLIGALLVCREGLAKEKQTQLHKEITLFAVPEDEKSMHSETKNSLMQDRD 240

Qy 241 AASARAPKGTAVNGYVNRSLPGLIGCRKSVYHVGWITPEVHSITPLSGHTFLVANH 300

Db 241 AASARAPKGTAVNGYVNRSLPGLIGCRKSVYHVGWITPEVHSITPLSGHTFLVANH 300

Qy 301 ROASLEISPIITFLTAQTLMDLQGLLSCHTSSHOHDGMEAYVKVDSCPEEPQLMKONE 360

Db 301 ROASLEISPIITFLTAQTLMDLQGLLSCHTSSHOHDGMEAYVKVDSCPEEPQLMKONE 360

Qy 361 EADYDDDLTDSMDVVRFDNDNSPFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420

Db 361 EADYDDDLTDSMDVVRFDNDNSPFTQIRSVAKKHPTWVHYIAAEEEDWDYAPLVLA 420

Qy 421 PDRSRYSKOYLNNGPQIRGRYKVRPMAYTDETKTREAIQHSGLIGPLLYGCVGDTL 480

Db 421 PDRSRYSKOYLNNGPQIRGRYKVRPMAYTDETKTREAIQHSGLIGPLLYGCVGDTL 480

Qy 481 LIIFKNQASRPVNTYFPHGTTDVRPLYSRRLPKGVKHLXDFPILPGEIPIKYKWTVVEDGP 540

Db 481 LIIFKNQASRPVNTYFPHGTTDVRPLYSRRLPKGVKHLXDFPILPGEIPIKYKWTVVEDGP 540

Qy 541 TKSDPRCLTRYSSFWNMRDLASGLIGPLLYCYKESVDQKQNMQRNVILFSVFDE 600

Db 541 TKSDPRCLTRYSSFWNMRDLASGLIGPLLYCYKESVDQKQNMQRNVILFSVFDE 600

Qy 601 NRSWYLTENIQFLNPAGVQLEDPEFQASNTWISNGVYFDSLOLSVCLHEVAYVYILS 660

Db 601 NRSWYLTENIQFLNPAGVQLEDPEFQASNTWISNGVYFDSLOLSVCLHEVAYVYILS 660

Qy 661 IGAQTDFLSVFSGYTFKKHWYEDTLTLPPFSGETVFMENPGLWILGCHNSDFNRG 720

Db 661 IGAQTDFLSVFSGYTFKKHWYEDTLTLPPFSGETVFMENPGLWILGCHNSDFNRG 720

Qy 721 MTALLKVSQCDKNTGYEDSYEDISAYLLSKNAIEPRSPFQNSRHPSTROKQFNATTI 780

Db 721 MTALLKVSQCDKNTGYEDSYEDISAYLLSKNAIEPRSPFQNSRHPSTROKQFNATTI 780

Qy 781 PENDIEKTDWFAHRTMPKIQNVSSDLMRLQSPTHGLSLSDLQZAKYETESDDPS 840

Db 781 PENDIEKTDWFAHRTMPKIQNVSSDLMRLQSPTHGLSLSDLQZAKYETESDDPS 840

Qy 841 PGAIIDNNSLSEMTHERPOLHHSQDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSST 900

Db 841 PGAIIDNNSLSEMTHERPOLHHSQDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSST 900

Qy 901 SNNLISTIPSDNLAAGTDNTSSILGPPSPMVHYDSOLDTTLPGKSSPLTSGGSPLSSEE 960

Db 901 SNNLISTIPSDNLAAGTDNTSSILGPPSPMVHYDSOLDTTLPGKSSPLTSGGSPLSSEE 960

Qy 961 NNDKLLLEGLMNSQSSWGKNTSVSTESGRLEFKGRAHGPALLTKONALFKVSIISLKTN 1020

Db 961 NNDKLLLEGLMNSQSSWGKNTSVSTESGRLEFKGRAHGPALLTKONALFKVSIISLKTN 1020

Qy 1021 KTSNNSATNRKTHIDGSPLLIENSPPVWQNILESDETFKKVTPLIHDMMLDKNATLRL 1080

Db 1021 KTSNNSATNRKTHIDGSPLLIENSPPVWQNILESDETFKKVTPLIHDMMLDKNATLRL 1080

Qy 1081 NMSNKTTSKQMEVQOKKEGIPPPDAQNPDSFFKMLFLPESARMIQTHGKNSLNSG 1140

Db 1081 NMSNKTTSKQMEVQOKKEGIPPPDAQNPDSFFKMLFLPESARMIQTHGKNSLNSG 1140

Qy 1141 QGPSPKQLVSLGPEKSVGQNFELSEKNKVVVGKEFTKVLKEMVPPSSRNFLTINLN 1200

Db 1141 QGPSPKQLVSLGPEKSVGQNFELSEKNKVVVGKEFTKVLKEMVPPSSRNFLTINLN 1200

Qy 1201 LHENNTNQKKEIEKKEKETTIOENVLPQIHTVGTGKXFMKNLFLSTRQNVESYD 1260

Db 1201 LHENNTNQKKEIEKKEKETTIOENVLPQIHTVGTGKXFMKNLFLSTRQNVESYD 1260

Qy 1261 GAYAPVTDQFSLNDSNTRKHTAHFSKGBEENLEGLNQIKQIVKCYACTRISPNT 1320

Db 1261 GAYAPVTDQFSLNDSNTRKHTAHFSKGBEENLEGLNQIKQIVKCYACTRISPNT 1320

Qy 1321 SQQNFVQSRKALQKRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

Db 1321 SQQNFVQSRKALQKRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE 1380

Qy 1381 KGAITQSPSLDCLTRSHSIPOANRSPPLIAKVSSFPPIRPIYLRVLFQDNSSHLPAASY 1440

Db 1381 KGAITQSPSLDCLTRSHSIPOANRSPPLIAKVSSFPPIRPIYLRVLFQDNSSHLPAASY 1440

Qy 1441 RKDQSGVQESSHFLQKAKKNNLSLAILTLEMTGQREVSGISGTSATNSVTYKVENTVLP 1500

Db 1441 RKDQSGVQESSHFLQKAKKNNLSLAILTLEMTGQREVSGISGTSATNSVTYKVENTVLP 1500

Qy 1501 KPDLPKTSGKVELLPKVHIYQKLPFTTETSNGSPGHLDLVEGSLLOQTGEGAIKWNEANRP 1560

Db 1501 KPDLPKTSGKVELLPKVHIYQKLPFTTETSNGSPGHLDLVEGSLLOQTGEGAIKWNEANRP 1560

Qy 1561 GKVPFLRVATSSAKTSPKLLDPLAWNHYGTQIPKEBWKQSKPEKTAFFKKKDTILSL 1620

Db 1561 GKVPFLRVATSSAKTSPKLLDPLAWNHYGTQIPKEBWKQSKPEKTAFFKKKDTILSL 1620

Qy 1621 NACESNHAIAINEGONKPELEVWAKQGRTERLCQNPVPLKSHOREITETTLQSDQEE 1680

Db 1621 NACESNHAIAINEGONKPELEVWAKQGRTERLCQNPVPLKSHOREITETTLQSDQEE 1680

Qy 1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSFOKKTTHYFIAAVERLDWYGMSSSHVLNRR 1740

Db 1681 IDYDDTISVEMKXEDFDIYDEDENQSPRSFOKKTTHYFIAAVERLDWYGMSSSHVLNRR 1740

Qy 1741 AQSQSVQFQKVVQFEETDGSFTQPLRGELNEHLGLGPYIRAEVDEINMVFERNQASR 1800
Db 1741 AQSQSVQFQKVVQFEETDGSFTQPLRGELNEHLGLGPYIRAEVDEINMVFERNQASR 1800
Qy 1801 PYSYSSLIISVEEORQGAERKQFVKNETKTYFWKVQHMAATKDEDFCKAWAYFSDV 1860
Db 1801 PYSYSSLIISVEEORQGAERKQFVKNETKTYFWKVQHMAATKDEDFCKAWAYFSDV 1860
Qy 1861 DLEKDVHSGLLGPIVLVCHTNTLNPAHGQVTVQBFALFFTFDFETKSYFTENNERCRA 1920
Db 1861 DLEKDVHSGLLGPIVLVCHTNTLNPAHGQVTVQBFALFFTFDFETKSYFTENNERCRA 1920
Qy 1921 PCNIQMEDPTFKENYRFHAINGYIMDTLPGVMAQDQIRWYLLSMGNSNENIHSIFSGH 1980
Db 1921 PCNIQMEDPTFKENYRFHAINGYIMDTLPGVMAQDQIRWYLLSMGNSNENIHSIFSGH 1980
Qy 1981 VFTVRKKEBEYKVALYNLYPGVFEVEMLPKAGIRWVECLIGEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEBEYKVALYNLYPGVFEVEMLPKAGIRWVECLIGEHLHAGMSTLFLVYSNKC 2040
Qy 2041 QTPGLMASGHIRDFOITASGOYGQWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Db 2041 QTPGLMASGHIRDFOITASGOYGQWAPKARLHYSGSINAWSTKEPFSWIKVDLLAPMII 2100
Qy 2101 HGIKTQGARQFSSLYISQFIIMYSLDGKKWQTVRGNSTGTLMVFPFGNVDSSGKHNIFN 2160
Db 2101 HGIKTQGARQFSSLYISQFIIMYSLDGKKWQTVRGNSTGTLMVFPFGNVDSSGKHNIFN 2160
Qy 2161 PPIIARYIRLHETHYSIRPSTLMEMLGCDLNSCPLGWSKASIDAOITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHETHYSIRPSTLMEMLGCDLNSCPLGWSKASIDAOITASSYFTNMFA 2220
Qy 2221 TWSPSKARLHQGRSNARPOVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHQGRSNARPOVNNPKEMQLQVDFQKTMKVTVGTTQGVKSLLTSMYVKEFL 2280
Qy 2281 ISSQDGHQWTLFPONGKVKVFGQNQDSFTPVVNSLDPLLTTRYLRHHPQSWVHQIALRM 2340
Db 2281 ISSQDGHQWTLFPONGKVKVFGQNQDSFTPVVNSLDPLLTTRYLRHHPQSWVHQIALRM 2340
Qy 2341 EVLGCEAODLY 2351
Db 2341 EVLGCEAODLY 2351

RESULT 3
US-10-133-907-4
; Sequence 4, Application US/10133907
; Publication No. US20030195223A1
; GENERAL INFORMATION:
; APPLICANT: Chien, Kenneth R
; APPLICANT: Hoshijima, Masahiko
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; TITLE OF INVENTION: Method to treat hemophilia by hepatic gene transfer of Factor VIII
; FILE REFERENCE: 6627-Pall170
; CURRENT APPLICATION NUMBER: US/10/133,907
; CURRENT FILING DATE: 2002-04-25
; PRIOR FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-907-4
Query Match 99.5%; Score 12399; DB 14; Length 2351;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYMOSDLGELFVDARFFPRVPKSPFN 60

Db 1 MQIELSTCFCLLRFCFSATRRYYLGAVELSWDYMOSDLGELFVDARFFPRVPKSPFN 60
Qy 61 TSVVYKXLTLEVEFTDHLFNIAKPRPPMMGLLGPITIOAEVYDVTVVITLKNWASHPVSLHAV 120
Db 61 TSVVYKXLTLEVEFTDHLFNIAKPRPPMMGLLGPITIOAEVYDVTVVITLKNWASHPVSLHAV 120
Qy 121 GVSYSKASEGAEXDDQTSOREKEDDKVFPQGSSTYVQVLKENGPMASDPLCLTYSYLSH 180
Db 121 GVSYSKASEGAEXDDQTSOREKEDDKVFPQGSSTYVQVLKENGPMASDPLCLTYSYLSH 180
Qy 181 VDLVKDNLNSGIGALVCREGSLAKETOTLHKFILLFAVFEDEGSKWSESTKSLMOORD 240
Db 181 VDLVKDNLNSGIGALVCREGSLAKETOTLHKFILLFAVFEDEGSKWSESTKSLMOORD 240
Qy 241 AASARAMPKMHVTNGYVNRSLPGLIGCHRKSVVHVHVIOMGTTPVHSGIFLEGHTFLVRNH 300
Db 241 AASARAMPKMHVTNGYVNRSLPGLIGCHRKSVVHVHVIOMGTTPVHSGIFLEGHTFLVRNH 300
Qy 301 ROASLEISPTITLTAQTLMLDGLQFLLSCHISSHOHGMAYVYVDSCEPQOLIMQNE 360
Db 301 ROASLEISPTITLTAQTLMLDGLQFLLSCHISSHOHGMAYVYVDSCEPQOLIMQNE 360
Qy 361 EADYDDDLTDSEMDVVRFDKDKSPSFIQIRSVAKKHPKTVHVIHAAEEDWDVAPLVLA 420
Db 361 EADYDDDLTDSEMDVVRFDKDKSPSFIQIRSVAKKHPKTVHVIHAAEEDWDVAPLVLA 420
Qy 421 PDDRSYKSQVLNNGPQIRGKVKVRFMAYTDETFKTREAIQHESGILGLLYGEVGDTL 480
Db 421 PDDRSYKSQVLNNGPQIRGKVKVRFMAYTDETFKTREAIQHESGILGLLYGEVGDTL 480
Qy 481 LIIFKNOASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFPKYKWTYVDEGP 540
Db 481 LIIFKNOASRPYNTYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFPKYKWTYVDEGP 540
Qy 541 TKSDPRLCTRYYSFVNMERDLASGLGPILLI CYKESVDOKGNOIMGDKRNVILFSVFDE 600
Db 541 TKSDPRLCTRYYSFVNMERDLASGLGPILLI CYKESVDOKGNOIMGDKRNVILFSVFDE 600
Qy 601 NRSWYLTENIORFLNPAGVQLEDPEFQASNMHSINGYVDFDSLQSLCYLHEVAYWILS 660
Db 601 NRSWYLTENIORFLNPAGVQLEDPEFQASNMHSINGYVDFDSLQSLCYLHEVAYWILS 660
Qy 661 IGAQTDLSLVFSGYTPKHKMAYEDTLTLPFSETVPMSENPCMLLGLCHNSDFNRG 720
Db 661 IGAQTDLSLVFSGYTPKHKMAYEDTLTLPFSETVPMSENPCMLLGLCHNSDFNRG 720
Qy 721 MTALLKVSSCDKNTGDYVEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTBQKQFNATTI 780
Db 721 MTALLKVSSCDKNTGDYVEDSYEDISAYLLSKNNAIEPRSPQNSRHPSTBQKQFNATTI 780
Qy 781 PENDIEKTPWFHRTMPKIQNVSSDMLMLLQSPTPHGLSLSDLOEAKYETFSDDPS 840
Db 781 PENDIEKTPWFHRTMPKIQNVSSDMLMLLQSPTPHGLSLSDLOEAKYETFSDDPS 840
Qy 841 PGADSNNSLSEMTHERPOLHSGDMVTPESGQLRNEKLGTTAATLKLKLPKVSST 900
Db 841 PGADSNNSLSEMTHERPOLHSGDMVTPESGQLRNEKLGTTAATLKLKLPKVSST 900
Qy 901 SNNLITSTIPDNLAAGTDNTSSLGPSPMPVHYDQSLDTTLTFGKSSPLTESGGFLSLEE 960
Db 901 SNNLITSTIPDNLAAGTDNTSSLGPSPMPVHYDQSLDTTLTFGKSSPLTESGGFLSLEE 960
Qy 961 NDSKLLBSGLMNSQESSWGKNVSTESGRFLFKRAHGPALLTKDNALFKVSI:SLKTN 1020
Db 961 NDSKLLBSGLMNSQESSWGKNVSTESGRFLFKRAHGPALLTKDNALFKVSI:SLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGSLLIENSPLYWQNTLESDTEFKVTPLIHORMKDNATLRL 1080
Db 1021 KTSNNSATNRKTHIDGSLLIENSPLYWQNTLESDTEFKVTPLIHORMKDNATLRL 1080
Qy 1081 NMSNKTTSKKNMNVQKKEGPI2PDAQNPDMSFFKMLFLPESARWQRTGKNSLNSG 1140
Db 1081 NMSNKTTSKKNMNVQKKEGPI2PDAQNPDMSFFKMLFLPESARWQRTGKNSLNSG 1140

```
QY 1141 QGSPKQVSLGPEKSEGVNQFVSEKNNVVGKGEFTKDVGLKEMVPPSSRNFLTNLDN 1200
Db 1141 QGSPKQVSLGPEKSEGVNQFVSEKNNVVGKGEFTKDVGLKEMVPPSSRNFLTNLDN 1200
QY 1201 LHENNTNQKKEQEBIEKKETLIQENWVLPQIHVTGTGNFMKNLFLLLTRQNVGSGYD 1260
Db 1201 LHENNTNQKKEQEBIEKKETLIQENWVLPQIHVTGTGNFMKNLFLLLTRQNVGSGYD 1260
QY 1261 GAYAPVLODRSLNDSNTRKKTAAHESKKEEENLEGLGNQTKQIVKCYACTRIISPNT 1320
Db 1261 GAYAPVLODRSLNDSNTRKKTAAHESKKEEENLEGLGNQTKQIVKCYACTRIISPNT 1320
QY 1321 SQQNFVTRSKRALQKRLPLEETEELKRIIVDTSTQWNNMKHLTPSTLTQIDYNEKE 1380
Db 1321 SQQNFVTRSKRALQKRLPLEETEELKRIIVDTSTQWNNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGALTQSPDLCDLTRSHSIPOANRSPPIAKVSSFPPIRPIYLRVLQONSSHLPAASY 1440
Db 1381 KGALTQSPDLCDLTRSHSIPOANRSPPIAKVSSFPPIRPIYLRVLQONSSHLPAASY 1440
QY 1441 RKXDSGVQESHHFLQGAKNLSLAITLTLEMTGDOREVSGLSGTSATNSVYKVENIVLP 1500
Db 1441 RKXDSGVQESHHFLQGAKNLSLAITLTLEMTGDOREVSGLSGTSATNSVYKVENIVLP 1500
QY 1501 KPDLPKTSGKVLLPKVHIYQKDLFPPTETSNQSPGHLDLVEGSLLOGTGEGAIKWNEANRP 1560
Db 1501 KPDLPKTSGKVLLPKVHIYQKDLFPPTETSNQSPGHLDLVEGSLLOGTGEGAIKWNEANRP 1560
QY 1561 GKVPFLVATESSAKTSPKLLDPLAWNHYGTQPKKEWKSQESPKPTAFKKXDTILSL 1620
Db 1561 GKVPFLVATESSAKTSPKLLDPLAWNHYGTQPKKEWKSQESPKPTAFKKXDTILSL 1620
QY 1621 NACSNHAIAINQGNKPELEVWAKQGRTERLCSQNPVVLKQHORBITRTILQSQEE 1680
Db 1621 NACSNHAIAINQGNKPELEVWAKQGRTERLCSQNPVVLKQHORBITRTILQSQEE 1680
QY 1681 IDYDNTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDNTISVEMKKEDFDIYDEDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHVLNR 1740
QY 1741 AQSQSVQKXVQVEFTDQSGFTQPLRGELNEHLGLLGPYIRAEVDENIMVYTRNQASR 1800
Db 1741 AQSQSVQKXVQVEFTDQSGFTQPLRGELNEHLGLLGPYIRAEVDENIMVYTRNQASR 1800
QY 1801 PYSFYSLISYEDORQAEPRKPVKNETKTYFKVQHEMAPTKDECDCKAWAYFSDV 1860
Db 1801 PYSFYSLISYEDORQAEPRKPVKNETKTYFKVQHEMAPTKDECDCKAWAYFSDV 1860
QY 1861 DLEKDVHSLIGPLLVCHTNTLNPAHQSVTVQEFALFTTIFDETQSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLIGPLLVCHTNTLNPAHQSVTVQEFALFTTIFDETQSWYFTENMERNCR 1920
QY 1921 PCNTQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGH 1980
Db 1921 PCNTQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRWYLLSMGSENIHSHFSGH 1980
QY 1981 VFTVRKKEEYKVALYNLYPGVFEIVEMLPKSGAGIWRVECLICEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKVALYNLYPGVFEIVEMLPKSGAGIWRVECLICEHLHAGMSTLFLVYSNKC 2040
QY 2041 QTPCLMGASHIRDFOITASGOYQWAPKARLHYSGS-NAWSTKBPFPFWIKVDLLAPMII 2100
Db 2041 QTPCLMGASHIRDFOITASGOYQWAPKARLHYSGS-NAWSTKBPFPFWIKVDLLAPMII 2100
QY 2101 HGKITQGARQKESLYISOFIIMYSLDGKKQTYRGNSTGTLMWFFGNVDSSGIXKHNFN 2160
Db 2101 HGKITQGARQKESLYISOFIIMYSLDGKKQTYRGNSTGTLMWFFGNVDSSGIXKHNFN 2160
QY 2161 PPIIARYIRLHPTHYSIRSLRMELMGCDLNSCSMPLGMESKAIQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPTHYSIRSLRMELMGCDLNSCSMPLGMESKAIQITASSYFTNMFA 2220
```

RESULT 4

```
US-10-411-037-30
; Sequence 30, Application US/10411037
; Publication No. US20040043446A1
; GENERAL INFORMATION:
; APPLICANT: Neose Technologies, Inc.
; APPLICANT: Defress, Shawn
; APPLICANT: Zopf, David
; APPLICANT: Bayer, Robert
; APPLICANT: Hakes, David
; APPLICANT: Chen, Xi
; APPLICANT: Rowe, Caryn
; TITLE OF INVENTION: ALPHA GALACTOSIDASE A: REMODELING AND GLYCOCONJUGATION OF ALPHA
; TITLE OF INVENTION: GALACTOSIDASE A
; FILE REFERENCE: 040853-01-5082
; CURRENT APPLICATION NUMBER: US/10/411,037
; CURRENT FILING DATE: 2003-04-09
; PRIOR APPLICATION NUMBER: US 60/328,523
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/344,692
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/387,292
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US 60/391,777
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/396,594
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US 60/404,249
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: US 60/407,527
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; TYPE: PRT
; LENGTH: 2351
; ORGANISM: Homo sapiens
US-10-411-037-30
```

Query Match 99,88; Score 12390; DB 12; Length 2351;

Best Local Similarity 99,88; Preg. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MQIELSTCFPLCLLRFCFSATRRYYLGAVELSWDYMOSDGLGELPVDARFPFRVPSPPFN 60
Db 1 MQIELSTCFPLCLLRFCFSATRRYYLGAVELSWDYMOSDGLGELPVDARFPFRVPSPPFN 60
QY 61 TSVVYKKTILFVEFTHDLFNIAKPPPPKMGILLGPTIQAEVYDTVTITLKNMASHPVSLHAV 120
Db 61 TSVVYKKTILFVEFTHDLFNIAKPPPPKMGILLGPTIQAEVYDTVTITLKNMASHPVSLHAV 120
QY 121 GVSYWKASEGAEYDDQTSQREKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLSH 180
Db 121 GVSYWKASEGAEYDDQTSQREKEDKVPFGGSHYVWQVLKENGPMASDPLCLTYSLSH 180
QY 181 VDLVKDNLNSGLIGALLVCREGSLAKBKTQTILHKTILFAVFEDEGKSWHSETKNSLMQORD 240
Db 181 VDLVKDNLNSGLIGALLVCREGSLAKBKTQTILHKTILFAVFEDEGKSWHSETKNSLMQORD 240
QY 241 AASRANPKKETHVANGYVNRSLPGLIGCHRKSVYWHVIGMTTPEVHSIFLEGHTFLVRNH 300
```

Db 241 AASARAWKMETUNGVYVRSI.PGLIGCHRKSVYVWVLCMGMTTPEVHSIFLBGHFTLVRNH 300
Qy 301 ROASLEISPTITFLTAQTLMLD.GOFLLSCHISSHOHDGMEAYVVDSCPEBPOLIMKNE 360
Db 301 ROASLEISPTITFLTAQTLMLD.GOFLLSCHISSHOHDGMEAYVVDSCPEBPOLIMKNE 360
Qy 361 BAEDYDDDLTDSEMDVVRFDNDSPSFTQIRSVAKKPKTWVHYIAABEEDWDVAPLVLA 420
Db 361 BAEDYDDDLTDSEMDVVRFDNDSPSFTQIRSVAKKPKTWVHYIAABEEDWDVAPLVLA 420
Qy 421 PDCRSYKQYLMMNGPQRIGRYKKVRFWAYTDETFTKTREAIQHSBGILGPLLYGEVGDTL 480
Db 421 PDCRSYKQYLMMNGPQRIGRYKKVRFWAYTDETFTKTREAIQHSBGILGPLLYGEVGDTL 480
Qy 481 LIIPKQASRPYNTYPHGIDTVRPLYSRRLPKGVKHLKDFPILPCELPFKYKWTVTVDGP 540
Db 481 LIIPKQASRPYNTYPHGIDTVRPLYSRRLPKGVKHLKDFPILPCELPFKYKWTVTVDGP 540
Qy 541 TKSDPRCLTRYYSFVNMERDLASGLIPLICYKESVDQKGNQIMSDKRNVLFSVFDE 600
Db 541 TKSDPRCLTRYYSFVNMERDLASGLIPLICYKESVDQKGNQIMSDKRNVLFSVFDE 600
Qy 601 NRSWYLTENIQRFLPNPAGVQLEDPEFOASNMHSINGYVDFSQLSVCLHEVAYWYILS 660
Db 601 NRSWYLTENIQRFLPNPAGVQLEDPEFOASNMHSINGYVDFSQLSVCLHEVAYWYILS 660
Qy 661 IGAOTDFLVSFPGVYTFKHKWYEDTLTLFPFSGETVFMSENPCGLWILGCHNSDFRNKG 720
Db 661 IGAOTDFLVSFPGVYTFKHKWYEDTLTLFPFSGETVFMSENPCGLWILGCHNSDFRNKG 720
Qy 721 MTALLKVSCKDNTGDYVDESDYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
Db 721 MTALLKVSCKDNTGDYVDESDYEDISAYLLSKNNAIEPRSPQNSRHPSTKQKFNATTI 780
Qy 781 PENDIEKTDPAHRTMPKIQNVSSDLMLLRQSPTPHGLSLSDLOEAYETPDDPS 840
Db 781 PENDIEKTDPAHRTMPKIQNVSSDLMLLRQSPTPHGLSLSDLOEAYETPDDPS 840
Qy 841 PGALDSNNLSGEMTHFRPOLHSHGDMVTPSGLOLRNEXLGTAAATELXKLDPKVSST 900
Db 841 PGALDSNNLSGEMTHFRPOLHSHGDMVTPSGLOLRNEXLGTAAATELXKLDPKVSST 900
Qy 901 SNNLTSTIPSDNLAAGTDNTSSGLPSPMPVHYDSQDNTTLFGKSSPTESGGPLSLSEE 960
Db 901 SNNLTSTIPSDNLAAGTDNTSSGLPSPMPVHYDSQDNTTLFGKSSPTESGGPLSLSEE 960
Qy 961 NNDKSLLESGLMNSQESSGKNVSVSTESGRLPKGRAGHPALLFKDNALPKVYSILLKTN 1020
Db 961 NNDKSLLESGLMNSQESSGKNVSVSTESGRLPKGRAGHPALLFKDNALPKVYSILLKTN 1020
Qy 1021 KTSNNSATNRKTHIDGSPLLTENSPSVWQNILESDETEFKVTPLTHDRMLMDKXATLRL 1080
Db 1021 KTSNNSATNRKTHIDGSPLLTENSPSVWQNILESDETEFKVTPLTHDRMLMDKXATLRL 1080
Qy 1081 NHMSNKTSSKNMVMQKKEGPIPPDAQNPDMSEFFKMLFIPESARWIQRTHGKNSLNSG 1140
Db 1081 NHMSNKTSSKNMVMQKKEGPIPPDAQNPDMSEFFKMLFIPESARWIQRTHGKNSLNSG 1140
Qy 1141 CGPSPKQVLVSLGPKEKSVGQNFLEBKKNVVGKGEFTKDVGLKEMVFPSSRNLPJTNLDN 1200
Db 1141 CGPSPKQVLVSLGPKEKSVGQNFLEBKKNVVGKGEFTKDVGLKEMVFPSSRNLPJTNLDN 1200
Qy 1201 LHENNTNHOEKKIQBEIEKKEKTLIQENVVLPOIHTVTGPKNFMKNLFLLSITRONVEGSDY 1260
Db 1201 LHENNTNHOEKKIQBEIEKKEKTLIQENVVLPOIHTVTGPKNFMKNLFLLSITRONVEGSDY 1260
Qy 1261 GAYAPVLQDFRSLNDSTNRTKXHTAHFSKKGBEENLEGJCNQTKOIVEKYACTTRISNT 1320
Db 1261 GAYAPVLQDFRSLNDSTNRTKXHTAHFSKKGBEENLEGJCNQTKOIVEKYACTTRISNT 1320
Qy 1321 SQONFVTOESKRALKQFRLPLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKE 1380

RESULT 5

US-10-411-026-30

Db 1321 SQONFVTOESKRALKQFRLPLEETELEKRIIVDDTSTQWSKMKHLTPSTLTQIDYNEKE 1380
Qy 1381 KGALTQSPDLSDCTRSHSIFQANRSPPLIAKVSSFPISIRPIYITRVLFQONSSELPAASY 1440
Db 1381 KGALTQSPDLSDCTRSHSIFQANRSPPLIAKVSSFPISIRPIYITRVLFQONSSELPAASY 1440
Qy 1441 RKXDSGVQESSHETLQGAKKNNLSLAITLBMTOQREVSGSLGTSATNSVYKKVENTVLP 1500
Db 1441 RKXDSGVQESSHETLQGAKKNNLSLAITLBMTOQREVSGSLGTSATNSVYKKVENTVLP 1500
Qy 1501 KPDLPTKISGVKVELLPKVHIYQKOLPPTETNGSPGHLDLVEGSLLOCTEGAIKWEANRP 1560
Db 1501 KPDLPTKISGVKVELLPKVHIYQKOLPPTETNGSPGHLDLVEGSLLOCTEGAIKWEANRP 1560
Qy 1561 GKVPFLRVATESAKTSPSKLLDPLAWNHYGTQIPKEWKSQEKSPKTAFFKKDITLSL 1620
Db 1561 GKVPFLRVATESAKTSPSKLLDPLAWNHYGTQIPKEWKSQEKSPKTAFFKKDITLSL 1620
Qy 1621 NACESNHAIALINBQONKPEIETWAKOGRTERLCSQNPVPLKHHQREITRTTLOSQDEE 1680
Db 1621 NACESNHAIALINBQONKPEIETWAKOGRTERLCSQNPVPLKHHQREITRTTLOSQDEE 1680
Qy 1681 IDYDDTISVENKKEDPDIYDEDENCSPRSFQKTRHYFIAAVERLMDYGVSSSHVLNR 1740
Db 1681 IDYDDTISVENKKEDPDIYDEDENCSPRSFQKTRHYFIAAVERLMDYGVSSSHVLNR 1740
Qy 1741 AOSGSVPQFKVVFQERTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTPRNOASR 1800
Db 1741 AOSGSVPQFKVVFQERTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTPRNOASR 1800
Qy 1801 PYSYSSLIISYEEDQRCAGSPRKNPVKPNETKTVFWKVQHHMAFTKOFDCKAWAYFSDV 1860
Db 1801 PYSYSSLIISYEEDQRCAGSPRKNPVKPNETKTVFWKVQHHMAFTKOFDCKAWAYFSDV 1860
Qy 1861 DLEKDVHSGILGPLLCHTNTLNPAHGSQVTVQBFALFTTFIDETKSWYFTENNERCRA 1920
Db 1861 DLEKDVHSGILGPLLCHTNTLNPAHGSQVTVQBFALFTTFIDETKSWYFTENNERCRA 1920
Qy 1921 PCNTQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRIWWYLLSMGSENENIHSIFSGH 1980
Db 1921 PCNTQMEDPTFKENYRPHAINGYIMDTLPGLVMAQDQIRIWWYLLSMGSENENIHSIFSGH 1980
Qy 1981 VFTVRKKEEYKMALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKMALYNLYPGVFTVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKC 2040
Qy 2041 QTPLGMAHGHRDFOITASGOYGOWAPKLAFLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
Db 2041 QTPLGMAHGHRDFOITASGOYGOWAPKLAFLHYSGSINAWSTKEPFSWIKVDLLAPMI 2100
Qy 2101 HGIKTQARQKFSSLYISQFIIMYSLDGKKQTYRGNSTGTLMVFPFGNVDSGKHNIFN 2160
Db 2101 HGIKTQARQKFSSLYISQFIIMYSLDGKKQTYRGNSTGTLMVFPFGNVDSGKHNIFN 2160
Qy 2161 PPIIARVIRLPHYTSIRSTRIMELMCDLNSCMPLGMEKASISDAQITASSYFTNMPA 2220
Db 2161 PPIIARVIRLPHYTSIRSTRIMELMCDLNSCMPLGMEKASISDAQITASSYFTNMPA 2220
Qy 2221 TWSPSKARLHLQGSNAWRPOVNNPKWLQVDFOKTKVTVTTQGVKSLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLQGSNAWRPOVNNPKWLQVDFOKTKVTVTTQGVKSLTSMYVKEFL 2280
Qy 2281 ISSQDGHQWTLFTQNGKVKVFCQNDQSFTEPVVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
Db 2281 ISSQDGHQWTLFTQNGKVKVFCQNDQSFTEPVVNSLDPPLLTRYLRIHPQSWHQAIALRM 2340
Qy 2341 EVLGCBAQDLY 2351
Db 2341 EVLGCBAQDLY 2351

Sequence 30, Application US/10411026
Publication No. US20040063911A1
GENERAL INFORMATION:
APPLICANT: Neose Technologies, Inc.
APPLICANT: DeFrees, Shawn
APPLICANT: Zopf, David
APPLICANT: Bayer, Robert
APPLICANT: Hakes, David
APPLICANT: Chen, Xi
TITLE OF INVENTION: PROTEIN REMODELING METHODS AND PROTEINS/PEPTIDES PRODUCED BY THE
TITLE OF INVENTION: METHODS
FILE REFERENCE: G40853-01-5053
CURRENT APPLICATION NUMBER: US/10/411,026
CURRENT FILING DATE: 2003-04-09
PRIORITY APPLICATION NUMBER: US 60/328,523
PRIORITY FILING DATE: 2001-10-10
PRIORITY APPLICATION NUMBER: US 60/344,692
PRIORITY FILING DATE: 2001-10-19
PRIORITY APPLICATION NUMBER: US 60/387,292
PRIORITY FILING DATE: 2002-06-07
PRIORITY APPLICATION NUMBER: US 60/391,777
PRIORITY FILING DATE: 2002-06-25
PRIORITY APPLICATION NUMBER: US 60/396,594
PRIORITY FILING DATE: 2002-07-17
PRIORITY APPLICATION NUMBER: US 60/404,249
PRIORITY FILING DATE: 2002-08-16
PRIORITY APPLICATION NUMBER: US 60/407,527
PRIORITY FILING DATE: 2002-08-28
NUMBER OF SEQ ID NOS: 75
SOFTWARE: PatentIn version 3.2
SEQ ID NO 30
LENGTH: 2351
TYPE: prt
ORGANISM: Homo sapiens
US-10-411-026-30

Query Match 99.8%; Score 12390; DB 12; Length 2351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2347; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 MQELSTCFCLLRFCFSATRYVGLGAVLSWDYMSDGLGELPVDARFPRVPEKSPFN 60
Db 1 MQELSTCFCLLRFCFSATRYVGLGAVLSWDYMSDGLGELPVDARFPRVPEKSPFN 60
QY 61 TSVYKKTLEVEFDHLFNIAKRPDPMMGLLGPITQAEVDTVVITLKNMASHPVSLHAV 120
Db 61 TSVYKKTLEVEFDHLFNIAKRPDPMMGLLGPITQAEVDTVVITLKNMASHPVSLHAV 120
QY 121 GVSYWKASEGAEDDQTSQREKEDKVPFGSHTYVQVILKENGPMASDFLCITYSLH 180
Db 121 GVSYWKASEGAEDDQTSQREKEDKVPFGSHTYVQVILKENGPMASDFLCITYSLH 180
QY 181 VDLVKDNLNSGLIGALLVCRGSLAKETQTLHKFILLFAVDFDEKSWHSETKNSLMODRD 240
Db 181 VDLVKDNLNSGLIGALLVCRGSLAKETQTLHKFILLFAVDFDEKSWHSETKNSLMODRD 240
QY 241 AAGARAWPKMHTVNGYVNRSLPGLGCHRXSVYVHWIEMGTTPPEVHSIFLEGHTFLVRNH 300
Db 241 AAGARAWPKMHTVNGYVNRSLPGLGCHRXSVYVHWIEMGTTPPEVHSIFLEGHTFLVRNH 300
QY 301 ROASLEISPIITELTAQTLMDLQGLFSLCHSSHQHGMENYKVVDSCPEEPQILMKNE 360
Db 301 ROASLEISPIITELTAQTLMDLQGLFSLCHSSHQHGMENYKVVDSCPEEPQILMKNE 360
QY 361 EADYDDDLTDSMDVVRFDNDSPSTQIRSVAKKPKTWVHWIAAEEEDWDVAPLVLA 420
Db 361 EADYDDDLTDSMDVVRFDNDSPSTQIRSVAKKPKTWVHWIAAEEEDWDVAPLVLA 420
QY 421 PDDRYSKQVYNNQPQRIGRKYKVRPMAYTDEFTKTRAIQHESGILGLLYGEVGDTL 480
Db 421 PDDRYSKQVYNNQPQRIGRKYKVRPMAYTDEFTKTRAIQHESGILGLLYGEVGDTL 480
QY 481 LIIFNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFKYKWTVTVEDGP 540

Db 481 LIIFNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDPILPGEIFKYKWTVTVEDGP 540
QY 541 TKSDDPECLTRYSSFFVNMERDLASGLIGPLLICYKESVDQKGNQIMSDKRNVLFSVFDE 600
Db 541 TKSDDPECLTRYSSFFVNMERDLASGLIGPLLICYKESVDQKGNQIMSDKRNVLFSVFDE 600
QY 601 NRSWYLTENIORFLPNPAGVQLEDPEFQASNMHSINGVYFDSLQSLVCLHEVAYWYLS 660
Db 601 NRSWYLTENIORFLPNPAGVQLEDPEFQASNMHSINGVYFDSLQSLVCLHEVAYWYLS 660
QY 661 IGAQTDFLSVFFSGYTFKHMVYEDTLTLPFSGGTVMFMSMENPGLWILGCHNSDPRNRG 720
Db 661 IGAQTDFLSVFFSGYTFKHMVYEDTLTLPFSGGTVMFMSMENPGLWILGCHNSDPRNRG 720
QY 721 MTALLKVVSSCDKNTGYEDSYEDI SAYLLSKNATEPRSFNSRHPSTRQKQFNATTI 780
Db 721 MTALLKVVSSCDKNTGYEDSYEDI SAYLLSKNATEPRSFNSRHPSTRQKQFNATTI 780
QY 781 PENDIEKTDPFWAHRTMPMKIQNVSSDMLMLLRQSPTEPHGLSLDLQBAKYETFSDDPS 840
Db 781 PENDIEKTDPFWAHRTMPMKIQNVSSDMLMLLRQSPTEPHGLSLDLQBAKYETFSDDPS 840
QY 841 PGAIDSNNSLSEMTHERPOLHHSGLDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSST 900
Db 841 PGAIDSNNSLSEMTHERPOLHHSGLDMVFTPESGQLRLNEKLGTTAATELKKLDFKVSST 900
QY 901 SNNLISTIPSDNLAACTDNTSSLGPPSPMVHYDQDQDITLFGKSSPLTESGGPLSLEE 960
Db 901 SNNLISTIPSDNLAACTDNTSSLGPPSPMVHYDQDQDITLFGKSSPLTESGGPLSLEE 960
QY 961 NNDKLLSGLMNSQSSSWGKNVSSSTESGRLFXGKRAHPALLTKNALFKVSIILLKTN 1020
Db 961 NNDKLLSGLMNSQSSSWGKNVSSSTESGRLFXGKRAHPALLTKNALFKVSIILLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPSILLIENSPSWQNLIESDTEFKKVTPLIHDRMLMDKNAALRL 1080
Db 1021 KTSNNSATNRKTHIDGPSILLIENSPSWQNLIESDTEFKKVTPLIHDRMLMDKNAALRL 1080
QY 1081 NMSNKTTSKNNMWWQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
Db 1081 NMSNKTTSKNNMWWQOKKEGPIPPDAQNPDMSFFKMLFLPESARWIQTHGKNSLNSG 1140
QY 1141 QGSPXQLVSLGPEKVEGQNFLEKKNVYVVGGEFTKDVGLAKEMVPPSSRNLFNTNLDN 1200
Db 1141 QGSPXQLVSLGPEKVEGQNFLEKKNVYVVGGEFTKDVGLAKEMVPPSSRNLFNTNLDN 1200
QY 1201 LHENNTNHOEKKTOEIEKKEKTLIOENVLPQIHTVGTGKNFMKNLFLSTRQNVGSGYD 1260
Db 1201 LHENNTNHOEKKTOEIEKKEKTLIOENVLPQIHTVGTGKNFMKNLFLSTRQNVGSGYD 1260
QY 1261 GAYAPVLQDFRSLNDSNTNRKTHAFHSKKGBEENLEGLGNQTKQIVEKYACTTRISPN 1320
Db 1261 GAYAPVLQDFRSLNDSNTNRKTHAFHSKKGBEENLEGLGNQTKQIVEKYACTTRISPN 1320
QY 1321 SQNFVQSRKALKQFRPLETELEKRIIIVDDTSTQWSKMKHILTPSTLTQIDYNEKE 1380
Db 1321 SQNFVQSRKALKQFRPLETELEKRIIIVDDTSTQWSKMKHILTPSTLTQIDYNEKE 1380
QY 1381 KGAITOSPLSDCLTRSHSIPOANRSLPTAKVSSPFSIRPIYLTRVLFQDNSSHLPAASY 1440
Db 1381 KGAITOSPLSDCLTRSHSIPOANRSLPTAKVSSPFSIRPIYLTRVLFQDNSSHLPAASY 1440
QY 1441 RKXDSGVQSSSHFLQGAKKNNLSAILITLMTDQREVGSLGTSATNSVTYKKEVENTVLP 1500
Db 1441 RKXDSGVQSSSHFLQGAKKNNLSAILITLMTDQREVGSLGTSATNSVTYKKEVENTVLP 1500
QY 1501 KPDLPXTSGKVELLPKVHIYQKDLFPTETSNQSGPHLDLVEGSLLOQTEGAIKWNEANRP 1560
Db 1501 KPDLPXTSGKVELLPKVHIYQKDLFPTETSNQSGPHLDLVEGSLLOQTEGAIKWNEANRP 1560
QY 1561 GKYPFLARVATESAKTPEKLLDPLANDNHYGTQIPKEEWSQESPEKTAFAKKDITLSL 1620

Db 1561 GKVPFLRVATSSAKTPSKLLDPLAWDHYHQVLPKEEWKQSEKSPKATPKKXDTILSL 1620
Qy 1621 NACESNHAIAINEGONKPELEVWAKOGRTERLCSQNPVLKXHQRIITRTTLQSDOE 1680
Db 1621 NACESNHAIAINEGONKPELEVWAKOGRTERLCSQNPVLKXHQRIITRTTLQSDOE 1680
Qy 1681 IDYDDTISVENKKEDFDYDDBENOSPRSFOKTRHYFAAVERLWDYGMSSSPHVLNR 1740
Db 1681 IDYDDTISVENKKEDFDYDDBENOSPRSFOKTRHYFAAVERLWDYGMSSSPHVLNR 1740
Qy 1741 AQSGSVPOFKVWFOEFTDGSFTQPLYGEINEHGLIGLPVIRAEVDNIMVTENQASR 1800
Db 1741 AQSGSVPOFKVWFOEFTDGSFTQPLYGEINEHGLIGLPVIRAEVDNIMVTENQASR 1800
Qy 1801 PYSFYSSLIISVEEDQOQAEPRKQFVKNEKTYFWKQVQHWMAPTKDFDCKAWAFSDV 1860
Db 1801 PYSFYSSLIISVEEDQOQAEPRKQFVKNEKTYFWKQVQHWMAPTKDFDCKAWAFSDV 1860
Qy 1861 DLEKDVHSLGILPILVCHTNILNPAHQVQVVBALFETTFDETKSWYFTENMERNCR 1920
Db 1861 DLEKDVHSLGILPILVCHTNILNPAHQVQVVBALFETTFDETKSWYFTENMERNCR 1920
Qy 1921 PCNQMEDPTKENTRFRFAINGYIMDTLPLGLVMAQDQIRWVLLSMGSENIHSHFSGH 1980
Db 1921 PCNQMEDPTKENTRFRFAINGYIMDTLPLGLVMAQDQIRWVLLSMGSENIHSHFSGH 1980
Qy 1981 VFTVRKKEEYKVALNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
Db 1981 VFTVRKKEEYKVALNLYPGVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKC 2040
Qy 2041 QTPLGMAHGHRDFOITASQGYQWAPKARLHYSGSINAMSTKPFPSWIKVDLLAPMII 2100
Db 2041 QTPLGMAHGHRDFOITASQGYQWAPKARLHYSGSINAMSTKPFPSWIKVDLLAPMII 2100
Qy 2101 HGKTCGARQKFPSSLIYISQFIIMYSLDGKQWOTYRGNSTGTLMVFGVNDSSGIXHNI 2160
Db 2101 HGKTCGARQKFPSSLIYISQFIIMYSLDGKQWOTYRGNSTGTLMVFGVNDSSGIXHNI 2160
Qy 2161 PPIIARYIRLHPTHYISIRSLAMELGMCDLNSCMLPGMESKAISDAQITASSYFTNMFA 2220
Db 2161 PPIIARYIRLHPTHYISIRSLAMELGMCDLNSCMLPGMESKAISDAQITASSYFTNMFA 2220
Qy 2221 TWSPSKARLHLOGRANRPPQNNKEMLOVDFOKTMKVGTTCGVKSLLTSMYVKEFL 2280
Db 2221 TWSPSKARLHLOGRANRPPQNNKEMLOVDFOKTMKVGTTCGVKSLLTSMYVKEFL 2280
Qy 2281 ISSSDQGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPPILLTRYLRIRHPQSWVHQIALRM 2340
Db 2281 ISSSDQGHQWTLFFQNGKVKVFCGNQDSFTPVVNSLDPPILLTRYLRIRHPQSWVHQIALRM 2340
Qy 2341 EVLGCEAODLY 2351
Db 2341 EVLGCEAODLY 2351

RESULT 6
US-10-360-101-229
; Sequence 229, Application US/10360.01
; Publication No. US20040009550A1
; GENERAL INFORMATION:
; APPLICANT: Moll, Gert N.
; APPLICANT: Leenhouts, Cornelis J.
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way
; FILE REFERENCE: 2183-5673
; CURRENT APPLICATION NUMBER: US/10/360,101
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: EP 02077060.8
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 309
; SOFTWARE: Patent-In version 3.1
; SEQ ID NO 229
; LENGTH: 2332
; TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: sequence of factor VIII
US-10-360-101-229
Query Match 99.0%; Score 12295; DB 15; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2329; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 20 ATRRYIYGAVLSWDYMQSDLGELPVDARPPRPVKPFPFNTSVYVKKTLFVBTFTDHLFN 79
Db 1 ATRRYIYGAVLSWDYMQSDLGELPVDARPPRPVKPFPFNTSVYVKKTLFVBTFTDHLFN 60
Qy 80 IAKRPPRMWGLGPTIQAQVYDVTYVITLKNMASHPVSLHVGVSYWKASGAEDDQTSQ 139
Db 61 IAKRPPRMWGLGPTIQAQVYDVTYVITLKNMASHPVSLHVGVSYWKASGAEDDQTSQ 120
Qy 140 REKEDDKVFPGGSHTYVWQVJLKNPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFPGGSHTYVWQVJLKNPMASDPLCLTYSLSHVDLVKDLNSGLIGALLVCR 180
Qy 200 EGSIAKEKTOHLKRPILLFAVVDGKSWHSETKSLMQDRDAASARAWPKMHTVNGYVR 259
Db 181 EGSIAKEKTOHLKRPILLFAVVDGKSWHSETKSLMQDRDAASARAWPKMHTVNGYVR 240
Qy 260 SLPLGIGCHRXSVVWHYIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 319
Db 241 SLPLGIGCHRXSVVWHYIGMGTTPVHSIFLEGHTFLVRNHRQASLEISPTFLTAQTL 300
Qy 320 MDLQQLFLLCHISHQHDGMEAYKVDSCPEEPQLIKNNEAEEDYDDLTDSMDVVRP 379
Db 301 MDLQQLFLLCHISHQHDGMEAYKVDSCPEEPQLIKNNEAEEDYDDLTDSMDVVRP 360
Qy 380 DDDNSPSFIQIRSAKKEPKTWVHYIAAEEDWDYAPLVAPDDRSYKSYQLNNGPQIRG 439
Db 361 DDDNSPSFIQIRSAKKEPKTWVHYIAAEEDWDYAPLVAPDDRSYKSYQLNNGPQIRG 420
Qy 440 RYKVKVRPMAYDTDTFTKTREAIQHESGILGPLLYGEVGDITLLIFPKQASRPYNIYPHI 499
Db 421 RYKVKVRPMAYDTDTFTKTREAIQHESGILGPLLYGEVGDITLLIFPKQASRPYNIYPHI 480
Qy 500 TDVPLYSRRIPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTYYSFVNME 559
Db 481 TDVPLYSRRIPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTYYSFVNME 540
Qy 560 RDLASGLIGPLLI CYKESVDQGNQIMSDKNVILFSVFDENRSMYLTENIQRFLPNAG 619
Db 541 RDLASGLIGPLLI CYKESVDQGNQIMSDKNVILFSVFDENRSMYLTENIQRFLPNAG 600
Qy 620 VQLEDPEFQASINIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 679
Db 601 VQLEDPEFQASINIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDPLSVFFSGYTFKH 660
Qy 680 KXVYEDTLTLPPFSGETVFMSENPGLWILCHNSDDRENRMGTALLKVSSCDKNTGDYIE 739
Db 661 KXVYEDTLTLPPFSGETVFMSENPGLWILCHNSDDRENRMGTALLKVSSCDKNTGDYIE 720
Qy 740 DSYEDISAYLLSKKNAIEPRFSQNSRHPSTRQKFNAITTPENDIEKTDWFARHTMP 799
Db 721 DSYEDISAYLLSKKNAIEPRFSQNSRHPSTRQKFNAITTPENDIEKTDWFARHTMP 780
Qy 800 KIQNVSSDILLMLLRQSPTPHGLSLSDLOEAKYTFSDPSGAIIDSNNSLSETHFRPQ 859
Db 781 KIQNVSSDILLMLLRQSPTPHGLSLSDLOEAKYTFSDPSGAIIDSNNSLSETHFRPQ 840
Qy 860 LHHSGDVYFTPESGLOLRNKLGTATATELKKLDKFKVSSSTNNLSTIPSDNLAAGTDN 919
Db 841 LHHSGDVYFTPESGLOLRNKLGTATATELKKLDKFKVSSSTNNLSTIPSDNLAAGTDN 900
Qy 920 TSSLGPPSPMPVHYTSQDITTLFGKKSPLTBSGGPLSLSEENNDKSLLESGLMNSQESSW 979
Db 901 TSSLGPPSPMPVHYTSQDITTLFGKKSPLTBSGGPLSLSEENNDKSLLESGLMNSQESSW 960

QY 960 GKNVSTESGRLFKGKRAHGPAALLTKDNALFKVSISSLKTKNTKNSNATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRLFKGKRAHGPAALLTKDNALFKVSISSLKTKNTKNSNATNRKTHIDGPSL 1020
QY 1040 LIENSPTSVMQILLESDBTFFKQVTPLIHDMMLMDKNATLRINEMSNKTTSSKNMEMVQOK 1099
Db 1021 LIENSPTSVMQILLESDBTFFKQVTPLIHDMMLMDKNATLRINEMSNKTTSSKNMEMVQOK 1080
QY 1100 KEGPIPPDAQNPDMSEFFKMLFLPESARWQTHGKNSNSGOGPSPKOLVSLGPEKSVEG 1159
Db 1081 KEGPIPPDAQNPDMSEFFKMLFLPESARWQTHGKNSNSGOGPSPKOLVSLGPEKSVEG 1140
QY 1160 QNFLSEKXVVVGKGEFTKDVGLKZMVPSSRNLFITNLDNLHNNHNBKKEQEEIEK 1219
Db 1141 QNFLSEKXVVVGKGEFTKDVGLKZMVPSSRNLFITNLDNLHNNHNBKKEQEEIEK 1200
QY 1220 KETLIOENVLPQIHVTGTFKMFNMLFLLSRQNVESYDCAVPLQDPRSLNDSNTR 1279
Db 1201 KETLIOENVLPQIHVTGTFKMFNMLFLLSRQNVESYDCAVPLQDPRSLNDSNTR 1260
QY 1280 TKKHTAHFSKKEBENLEGLNQTKQIVKQVACTTRISPNSTSQNFVTRQKRALKQPRL 1339
Db 1261 TKKHTAHFSKKEBENLEGLNQTKQIVKQVACTTRISPNSTSQNFVTRQKRALKQPRL 1320
QY 1340 PLETELEKRIIIVDDTSTQWKNMKHLPSTLTQIDYNEKXKGAITQSPISDCULTRSHI 1399
Db 1321 PLETELEKRIIIVDDTSTQWKNMKHLPSTLTQIDYNEKXKGAITQSPISDCULTRSHI 1380
QY 1400 POANRSPPIAKVSFSPRIPIYTRVLFOQNSHLPAASRYKXKDSGVQESSHFLQGA 1459
Db 1381 POANRSPPIAKVSFSPRIPIYTRVLFOQNSHLPAASRYKXKDSGVQESSHFLQGA 1440
QY 1460 NKLSLAILTLEMTGQOREVSGISATNSVTVKVENTVLKPDLPKTSKGVLELLPKVHI 1519
Db 1441 NKLSLAILTLEMTGQOREVSGISATNSVTVKVENTVLKPDLPKTSKGVLELLPKVHI 1500
QY 1520 YOKOLFPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPKVPFLRVATESAKTPSK 1579
Db 1501 YOKOLFPTETSGSPGHLDLVEGSLLOQTEGAIKWNEANRPKVPFLRVATESAKTPSK 1560
QY 1580 LLDPLAWNHYGTQIPKEEWSQKSPKTAFFKDDTILSNACESHAIINAEQGNKP 1639
Db 1561 LLDPLAWNHYGTQIPKEEWSQKSPKTAFFKDDTILSNACESHAIINAEQGNKP 1620
QY 1640 BIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLQSDQEBIDYDDTISVMKKEDFDIY 1699
Db 1621 BIEVTWAKQGRTERLCSQNPVLRKHOREITRTTLQSDQEBIDYDDTISVMKKEDFDIY 1680
QY 1700 DEDENGSPRSFQKTRHYFIAAVERLWDYGNSSPHVLNRNAQSGSVPOPKVVPQSFDT 1759
Db 1681 DEDENGSPRSFQKTRHYFIAAVERLWDYGNSSPHVLNRNAQSGSVPOPKVVPQSFDT 1740
QY 1760 GSFTQPLRYGELNEHGLGPLYIRAEVEDNIMVTRFQNASRPFYSLSLISYEEDORQGA 1819
Db 1741 GSFTQPLRYGELNEHGLGPLYIRAEVEDNIMVTRFQNASRPFYSLSLISYEEDORQGA 1800
QY 1820 EPRNFVKNPNETKYFWKQVQHMAFTKDEPCKAWAYFSDVLEKDVHSGLIGPLLCHT 1879
Db 1801 EPRNFVKNPNETKYFWKQVQHMAFTKDEPCKAWAYFSDVLEKDVHSGLIGPLLCHT 1860
QY 1880 NTLNPAHGRQVTVQEFALFFTFIDETKSWYFTEMMERNCRAPCNIQWEDPTFKENYRFA 1939
Db 1861 NTLNPAHGRQVTVQEFALFFTFIDETKSWYFTEMMERNCRAPCNIQWEDPTFKENYRFA 1920
QY 1940 INGVIMDTPLGVVQAQDRIWYLLSNGSNENIHSIFSGHVFTVRKKEEYKALYNLYP 1999
Db 1921 INGVIMDTPLGVVQAQDRIWYLLSNGSNENIHSIFSGHVFTVRKKEEYKALYNLYP 1980
QY 2000 GVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKKQCTPLGNASGHIRDFOITAS 2059
Db 1981 GVFTVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKKQCTPLGNASGHIRDFOITAS 2040
QY 2060 GOYQOWAPKLARLHYSGSINAWSKEPSPWIKVDLLAPMIHGIKTQAROKFSSJISQ 2119

Db 2041 GOYQOWAPKLARLHYSGSINAWSKEPSPWIKVDLLAPMIHGIKTQAROKFSSJISQ 2100
QY 2120 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSSGFKNNIPNPPIIARYIRLHPHYSIRS 2179
Db 2101 FIIMYSLDGKQWQYRGNSTGTLMVFFGNVDSSGFKNNIPNPPIIARYIRLHPHYSIRS 2160
QY 2180 TLRVLMSCDNLNCSMPLGMSKALSDAQITASSVFTNNFATWSPSKARLHLOGRSNAR 2239
Db 2161 TLRVLMSCDNLNCSMPLGMSKALSDAQITASSVFTNNFATWSPSKARLHLOGRSNAR 2220
QY 2240 PQVNNPKEWLQVDQKTMKVTVTTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2299
Db 2221 PQVNNPKEWLQVDQKTMKVTVTTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKV 2280
QY 2300 KYFGQNDSPFPVNSLDPPLLTLYRLHRIHPOSVHQIALRMEVLGCEAQDLY 2351
Db 2281 KYFGQNDSPFPVNSLDPPLLTLYRLHRIHPOSVHQIALRMEVLGCEAQDLY 2332
RESULT 7
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; CURRENT FILING DATE: 2001-09-16
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-957-641-2
Query Match 98.9%; Score 12282; DB 9; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 20 ATRRYYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEPTHLEN 79
Db 1 ATRRYYLGAVELSDWYMQSDLGELPVDARPPRPVKSPFNTSVVYKTLFVEPTHLEN 60
QY 80 IAKPRPFWMLGLGPTIOAEVYDVITLKNWASHPVSLHAGVSYWKASGAEYDDOTSQ 139
Db 61 IAKPRPFWMLGLGPTIOAEVYDVITLKNWASHPVSLHAGVSYWKASGAEYDDOTSQ 120
QY 140 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLICALLVCR 199
Db 121 REKEDDKVFPQGSHTYVWQVLKENGPMASDPLCTYSYLSHVDLVKDLNSGLICALLVCR 180
QY 200 EGSLAKETQTLHKFILLFAVFBEGSKSWHSETKNSLMODRDAASARAWPKMHTVNGVNR 259
Db 181 EGSLAKETQTLHKFILLFAVFBEGSKSWHSETKNSLMODRDAASARAWPKMHTVNGVNR 240
QY 260 SLPLGLICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHOASLEISPIITFLTAOTLL 319
Db 241 SLPLGLICHRKSVVHVIGMTTPEVHSIFLEGHTFLVRNHOASLEISPIITFLTAOTLL 300
QY 320 MDLGQFLLSCHISQHDGMEAYVKVDSCEPEQLIMKNEEAEDYDDDLTDSMDVYRF 379
Db 301 MDLGQFLLSCHISQHDGMEAYVKVDSCEPEQLIMKNEEAEDYDDDLTDSMDVYRF 360
QY 380 DDNSPSFPIQIRSVAKKPKTWHTYIAAEEDWDYAPLVAPDDRSYKSYLNNGPORIG 439
Db 361 DDNSPSFPIQIRSVAKKPKTWHTYIAAEEDWDYAPLVAPDDRSYKSYLNNGPORIG 420

QY 440 RYKXVRFMAYTDETFKTRAEAIQHESGILGPLYGEVGDITLIIIFKNQASRPYNTYPHGI 499
DB 421 RYKXVRFMAYTDETFKTRAEAIQHESGILGPLYGEVGDITLIIIFKNQASRPYNTYPHGI 480
QY 500 TVDRPLYSHRLSGVXKHLKDFILPQEPKYPKWTVVDEGPKTSDBPRCLTRYYSFVNM 559
DB 481 TVDRPLYSHRLSGVXKHLKDFILPQEPKYPKWTVVDEGPKTSDBPRCLTRYYSFVNM 540
QY 560 RDLASGLIGPLIICYKESVDQXGNOIMSDKRVILFSPFDENSXVLTEIORFLPNPAG 619
DB 541 RDLASGLIGPLIICYKESVDQXGNOIMSDKRVILFSPFDENSXVLTEIORFLPNPAG 600
QY 620 VOLEPPEQASIMESINGYVFDLSQLSVCLHEVAYWYILSIGAOTDFLSVFSGYTFKH 679
DB 601 VOLEPPEQASIMESINGYVFDLSQLSVCLHEVAYWYILSIGAOTDFLSVFSGYTFKH 660
QY 680 KMYVEDTLTLFPFSGETVPMSENEGLATILGCHNSDFRNMGMTALLKVSSCDKNTGDYIE 739
DB 661 KMYVEDTLTLFPFSGETVPMSENEGLATILGCHNSDFRNMGMTALLKVSSCDKNTGDYIE 720
QY 740 DSYEDISAYLLSKNNAIPRSPSONSRHPSRQKOFNATITIPENDIEKTDPMWAEHRTMP 799
DB 721 DSYEDISAYLLSKNNAIPRSPSONSRHPSRQKOFNATITIPENDIEKTDPMWAEHRTMP 780
QY 800 KTONVSSDMLLRQSPHGLSDLSLQEAKEYTFSDPSPGCAIDSNNSLSEMTFRPQ 859
DB 781 KTONVSSDMLLRQSPHGLSDLSLQEAKEYTFSDPSPGCAIDSNNSLSEMTFRPQ 840
QY 860 LHSXGDMVFTPESGLOLRNEKLTGTAATELKLDPKVSSTSNNLSTIPSDNLAAGTDN 919
DB 841 LHSXGDMVFTPESGLOLRNEKLTGTAATELKLDPKVSSTSNNLSTIPSDNLAAGTDN 900
QY 920 TSSLGPPNMPVHYDQLDITLPGKSSPLTESGGPLSISENNDKLLSGLMNSQESSW 979
DB 901 TSSLGPPNMPVHYDQLDITLPGKSSPLTESGGPLSISENNDKLLSGLMNSQESSW 960
QY 980 GKNVSTSGRLFKGRAGPALLTKDNALFKVSISLLKTNKTSNNSATNRKTHIDGPSL 1039
DB 961 GKNVSTSGRLFKGRAGPALLTKDNALFKVSISLLKTNKTSNNSATNRKTHIDGPSL 1020
QY 1040 LIENSPPVQWQILESDTBFKVTPLIHRMLMDKNATAILNHSNKTTSKNMWMQOK 1099
DB 1021 LIENSPPVQWQILESDTBFKVTPLIHRMLMDKNATAILNHSNKTTSKNMWMQOK 1080
QY 1100 KEGPIPPDAQNPDXFFXOMLFLPESARMIQRTGKNSLNSGGPSPKQVLSLGPESKSV 1159
DB 1081 KEGPIPPDAQNPDXFFXOMLFLPESARMIQRTGKNSLNSGGPSPKQVLSLGPESKSV 1140
QY 1160 QNPLSEKNKVVKGEPFKDVGKEMVPESSRNLFITNLDNLHENTHNOBKKEOEIEK 1219
DB 1141 QNPLSEKNKVVKGEPFKDVGKEMVPESSRNLFITNLDNLHENTHNOBKKEOEIEK 1200
QY 1220 KETLITQENNVLPQIHTVTGKTNFMKNLFLSTRQNVESYGAYAPVLODRSLNDSNR 1279
DB 1201 KETLITQENNVLPQIHTVTGKTNFMKNLFLSTRQNVESYGAYAPVLODRSLNDSNR 1260
QY 1280 TKKHTAHSKKEENLEBGLNQTKQIWEKYACTTRISNTSQQNFVQSRKRALQPRL 1339
DB 1261 TKKHTAHSKKEENLEBGLNQTKQIWEKYACTTRISNTSQQNFVQSRKRALQPRL 1320
QY 1340 PLEETELEKRIIVDDTQWNSKMHLPSTLTQIDYNEKEKGAITQSPDCLTRSHSI 1399
DB 1321 PLEETELEKRIIVDDTQWNSKMHLPSTLTQIDYNEKEKGAITQSPDCLTRSHSI 1380
QY 1400 POANRSPPLIAKVSSFPPIRPIYLRVLFDQNSHLPAAVYRKDQSGVQESSHFLQGAKK 1459
DB 1381 POANRSPPLIAKVSSFPPIRPIYLRVLFDQNSHLPAAVYRKDQSGVQESSHFLQGAKK 1440
QY 1460 NNLSLAULTLEMTGQREVGSLGTSATNSVTYKVKVENTVLKPDLPKTSKGVELLPKVHI 1519
DB 1441 NNLSLAULTLEMTGQREVGSLGTSATNSVTYKVKVENTVLKPDLPKTSKGVELLPKVHI 1500

QY 1520 YOKOLFPPTETNGSGPHGLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTSPK 1579
DB 1501 YOKOLFPPTETNGSGPHGLDLVEGSLLOQTEGAIKWNEANRPGKVPFLRVATESSAKTSPK 1560
QY 1580 LLDPLAWDNHYGTQIPKEEWSQESKPEKTAFAKKDITLSLNACESNHAIAINEGQNK 1639
DB 1561 LLDPLAWDNHYGTQIPKEEWSQESKPEKTAFAKKDITLSLNACESNHAIAINEGQNK 1620
QY 1640 ELEVTAQKQTERLCSQNPVLKXHQBEITRTTLOSQOEIDIDYDDTISVEMKEDFDIY 1699
DB 1621 ELEVTAQKQTERLCSQNPVLKXHQBEITRTTLOSQOEIDIDYDDTISVEMKEDFDIY 1680
QY 1700 DEDENQSPRSQKKTIRHYFIAAVERLMDYGMSSPHVLNRNAQSGSVPOFKVWVQFETD 1759
DB 1681 DEDENQSPRSQKKTIRHYFIAAVERLMDYGMSSPHVLNRNAQSGSVPOFKVWVQFETD 1740
QY 1760 GSFTQPIYRGELNBHLGLLGPYISAEVEDNIMVTRNOASRPYFYSLSIYEBEDQROGA 1819
DB 1741 GSFTQPIYRGELNBHLGLLGPYISAEVEDNIMVTRNOASRPYFYSLSIYEBEDQROGA 1800
QY 1820 EPRKNFVKPNETKTYFKVOHMMAPTKEDEPCKKAWPDSVDLEKDVHSGLIGLILVCHT 1879
DB 1801 EPRKNFVKPNETKTYFKVOHMMAPTKEDEPCKKAWPDSVDLEKDVHSGLIGLILVCHT 1860
QY 1880 NTLAPAGRQVTVQEFALFETIFDETCSWYFTENMERNCRAPCNIOEMDPTFKENYRPHA 1939
DB 1861 NTLAPAGRQVTVQEFALFETIFDETCSWYFTENMERNCRAPCNIOEMDPTFKENYRPHA 1920
QY 1940 INGVIMDTPLGVLVNAQDQRIWYLLSNGSNENIHSIHPSGHVFTVRKKEEYKMALYNLYP 1999
DB 1921 INGVIMDTPLGVLVNAQDQRIWYLLSNGSNENIHSIHPSGHVFTVRKKEEYKMALYNLYP 1980
QY 2000 GVETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGMASCHIRPOFITAS 2059
DB 1981 GVETVEMLPKAGIWRVECLIGELHAGMSTLFLVYSNKCOTPLGMASCHIRPOFITAS 2040
QY 2060 GOYQWAPKLARLHYSGSINAWSTKEPPSWTKVLLAPMIHGIKTQAROKFSSLYISQ 2119
DB 2041 GOYQWAPKLARLHYSGSINAWSTKEPPSWTKVLLAPMIHGIKTQAROKFSSLYISQ 2100
QY 2120 FIIMYSLDGKWKQYRGNSTGTLNVFFGNVDSSGIKNINFPDPIIARYIRLHPHYSIRS 2179
DB 2101 FIIMYSLDGKWKQYRGNSTGTLNVFFGNVDSSGIKNINFPDPIIARYIRLHPHYSIRS 2160
QY 2180 TLRVLMGCDLNSCMLPGMBSKASDAQITASSYFTNMPATWSPSKARLHLOGRSNAWR 2239
DB 2161 TLRVLMGCDLNSCMLPGMBSKASDAQITASSYFTNMPATWSPSKARLHLOGRSNAWR 2220
QY 2240 POVNNPKWLOVDQKTKKVTGTQGVKSLLTSMYVKEFLISSQDGHQWTLFPQNGKV 2299
DB 2221 POVNNPKWLOVDQKTKKVTGTQGVKSLLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KVFGQNSDFTFVNSLDPPLTBYLRIHPQSWHQIALRMEVLGCEAQDLY 2351
DB 2281 KVFGQNSDFTFVNSLDPPLTBYLRIHPQSWHQIALRMEVLGCEAQDLY 2332

RESULT 8

US-10-187-319-2
; Sequence 2, Application US/10187319
; Publication No. US2003006878A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:

```

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/187,319
/ FILING DATE: 27-Aug-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 09/523,656
/ FILING DATE: 2000-03-10
/ APPLICATION NUMBER: US 09/037,601
/ FILING DATE: 1998-03-10
/ APPLICATION NUMBER: WO PCT/US97/11155
/ FILING DATE: 1997-06-26
/ APPLICATION NUMBER: US 06/670,707
/ FILING DATE: 1996-06-26
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Greenlee, Lorance L.
/ REGISTRATION NUMBER: 27,894
/ REFERENCE/DOCKET NUMBER: 75-95K
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 303/499-8080
/ TELEFAX: 303/499-8089
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2332 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: <Unknown>
/ MOLECULE TYPE: protein
/ HYPOTHEetical: YES
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: N-terminal
/ ORIGINAL SOURCE:
/ ORGANISM: Homo sapiens
/ TISSUE TYPE: Liver
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
/ US-10-187-319-2
/
/ Query Match 98.98; Score 12282; DB 14; Length 2332;
/ Best Local Similarity 99.88; Pred. No. 0;
/ Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
/
/ QY 20 ATRYILGAVELSDWYMQSDLGELPVDARPPRPVKSFPENTSVYKTLFVFTDHLFN 79
/ Db 1 ATRYILGAVELSDWYMQSDLGELPVDARPPRPVKSFPENTSVYKTLFVFTDHLFN 60
/
/ QY 80 IAKPRPPMGLLPTTQAEVYDVTWITLKNASHPVSLHAGVSYWKASGAEYDDQTSQ 139
/ Db 61 IAKPRPPMGLLPTTQAEVYDVTWITLKNASHPVSLHAGVSYWKASGAEYDDQTSQ 120
/
/ QY 140 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSYLSHVDLVXDLSGLLGALLVCR 199
/ Db 121 REKEDKVPFGSGHTYVQVLKENGPMASDPLCLTYSYLSHVDLVXDLSGLLGALLVCR 180
/
/ QY 200 EGSIAKKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPQHTWNGVYNR 259
/ Db 181 EGSIAKKTQTLHKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPQHTWNGVYNR 240
/
/ QY 260 SLPLGLGCHRSKYVWVIGMGTTPVHSIFLEGHTFIVNRHQAASLEISPTIFLTATLL 319
/ Db 241 SLPLGLGCHRSKYVWVIGMGTTPVHSIFLEGHTFIVNRHQAASLEISPTIFLTATLL 300
/
/ QY 320 MDLGFQLLSCHISHQHDGMEAYVKVDSCPEEPQLIMKNEEAEYDDDLTDSMDVVRP 379
/ Db 301 MDLGFQLLPCHISHQHDGMEAYVKVDSCPEEPQLRMKNEEAEYDDDLTDSMDVVRP 360
/
/ QY 380 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVAPDDRYSKSOYLNNGPQIRG 439
/ Db 361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEDWDYAPLVAPDDRYSKSOYLNNGPQIRG 420
/
/ QY 440 RYKVKRFMAYTDETPKTRAIQHESGILGPLLYGEVGDFTLLIIPKQASRPNIYPHGI 499

```

```

/ Db 421 RYKVKRFMAYTDETPKTRAIQHESGILGPLLYGEVGDFTLLIIPKQASRPNIYPHGI 480
/
/ QY 500 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEDGFTKSDPRCLTRYSSVFVME 559
/ Db 481 TDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKVTVTVEDGFTKSDPRCLTRYSSVFVME 540
/
/ QY 560 RDLASGLIGELLICYKESVDOKGNOIMSDKRNVLFSVFDENRSWYLTENIQFLNPAG 619
/ Db 541 RDLASGLIGELLICYKESVDQGNQIMSDKRNVLFSVFDENRSWYLTENIQFLNPAG 600
/
/ QY 620 VQLEDEPFOASNMHMSINGVYVDSLOL SVCLHEVAYWYILSIGAQTDFL SVFSPGYTFKH 679
/ Db 601 VQLEDEPFOASNMHMSINGVYVDSLOL SVCLHEVAYWYILSIGAQTDFL SVFSPGYTFKH 660
/
/ QY 680 KMVYEDTLILFPFSGETVFMENPGLWILGCHNSDPNRGMPTALLKVSSCDKXNTGYE 739
/ Db 661 KMVYEDTLILFPFSGETVFMENPGLWILGCHNSDPNRGMPTALLKVSSCDKXNTGYE 720
/
/ QY 740 DSYEDISAYLLSKNNAIEPRSPNSRHPSTRQKNATTPENDIEKTPWFARHTPMP 799
/ Db 721 DSYEDISAYLLSKNNAIEPRSPNSRHPSTRQKNATTPENDIEKTPWFARHTPMP 780
/
/ QY 800 KIQNVSSDILLMLRQSPTPHGLSLDLQAKYETFSDDPSGALDSNNLSSEWTHRPQ 859
/ Db 781 KIQNVSSDILLMLRQSPTPHGLSLDLQAKYETFSDDPSGALDSNNLSSEWTHRPQ 840
/
/ QY 860 LHSXGDMVFTPEGSLQRLNEKLGTTAAELKKLDFKVSTSNKLISTIPSDMLAAGTDN 919
/ Db 841 LHSXGDMVFTPEGSLQRLNEKLGTTAAELKKLDFKVSTSNKLISTIPSDMLAAGTDN 900
/
/ QY 920 TSSLGPPSMPVHYDSQDITTLFCKKSSPTESGGPI SLSENNDSKLLSEGLMNSQESSW 979
/ Db 901 TSSLGPPSMPVHYDSQDITTLFCKKSSPTESGGPI SLSENNDSKLLSEGLMNSQESSW 960
/
/ QY 980 GKNVSTESGRLPFKGRAHGPALLITKDNALFKVSI SLLTKNTSNNSATNRKTHIDGPSL 1039
/ Db 961 GKNVSTESGRLPFKGRAHGPALLITKDNALFKVSI SLLTKNTSNNSATNRKTHIDGPSL 1020
/
/ QY 1040 LIENSFVQWNLLESDTEPKVTPLIHDRMLMDKNATALLRNHMSKNTTSSKXNMVQOK 1099
/ Db 1021 LIENSFVQWNLLESDTEPKVTPLIHDRMLMDKNATALLRNHMSKNTTSSKXNMVQOK 1080
/
/ QY 1100 KEGPIPDACNPDMSFFKMLFLPESARWORTHGKNSLNSGGQSPSKQLVSLGPEKSVEG 1159
/ Db 1081 KEGPIPDACNPDMSFFKMLFLPESARWORTHGKNSLNSGGQSPSKQLVSLGPEKSVEG 1140
/
/ QY 1160 QNPLSEKNKVVGKGFTHKDVGLKEMVFPSSRNPLTLNLDNLHNNHNNHQQEKIOEIEK 1219
/ Db 1141 QNPLSEKNKVVGKGFTHKDVGLKEMVFPSSRNPLTLNLDNLHNNHNNHQQEKIOEIEK 1200
/
/ QY 1220 KETLIOENVVLQIHTVTGTNFMKNLFLLSVTRQNVGSDGAYAPVLODFSLNDSTNR 1279
/ Db 1201 KETLIOENVVLQIHTVTGTNFMKNLFLLSVTRQNVGSDGAYAPVLODFSLNDSTNR 1260
/
/ QY 1280 TKKHTAHFKKGBEENLEGLGNQTKOIVEKYACTTRISNTSQNFVTOQRKALQKFL 1339
/ Db 1261 TKKHTAHFKKGBEENLEGLGNQTKOIVEKYACTTRISNTSQNFVTOQRKALQKFL 1320
/
/ QY 1340 PLEETELEXRIIVDDTSTQWKNMKHLETSITQIDYNEKEKGAITQSPSLDCLTRSHSI 1399
/ Db 1321 PLEETELEXRIIVDDTSTQWKNMKHLETSITQIDYNEKEKGAITQSPSLDCLTRSHSI 1380
/
/ QY 1400 POANRSPDLIAKVSFPSPRIPIYLRVLFDQNSHLLPAASYFKKDSGVQESSHFLQCAKK 1459
/ Db 1381 POANRSPDLIAKVSFPSPRIPIYLRVLFDQNSHLLPAASYFKKDSGVQESSHFLQCAKK 1440
/
/ QY 1460 NNLISAILILEMTGQREYVGLGT SATNSVTYKVENTVLPKPDLPKXTSGKVELLPKVHI 1519
/ Db 1441 NNLISAILILEMTGQREYVGLGT SATNSVTYKVENTVLPKPDLPKXTSGKVELLPKVHI 1500
/
/ QY 1520 YQKDLFPPTFTSGSFEGLDLVEGSLQGTGEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1579

```

Db 1501 YQKLLFPTETNGSGCHLDLVEGSLLOCTEGAIKWNEANRPGKVPFLRVATESSAKTPSK 1560
QY 1580 LLDPLAWNHVGTQTPKBEWSQBSPEKTAFKKOTILSNACSNHAIINAGNQKP 1639
Db 1561 LLDPLAWNHVGTQTPKBEWSQBSPEKTAFKKOTILSNACSNHAIINAGNQKP 1620
QY 1640 ELEVTAQGRTERCSQNPVPLKHHQREIIRTTTQSQDEEIDYDDTISVEMKXEDFDIY 1699
Db 1621 ELEVTAQGRTERCSQNPVPLKHHQREIIRTTTQSQDEEIDYDDTISVEMKXEDFDIY 1680
QY 1700 DEDENQSRFQKTRHYPIAVERLWYGMSSPHVLRNRAQSGSVPOFKVVFQETD 1759
Db 1681 DEDENQSRFQKTRHYPIAVERLWYGMSSPHVLRNRAQSGSVPOFKVVFQETD 1740
QY 1760 CSFTQPLVRELNEHLGLGPIVIRAEVDNLMVTRNQASRPSYSLSIYSEEDOROGA 1819
Db 1741 CSFTQPLVRELNEHLGLGPIVIRAEVDNLMVTRNQASRPSYSLSIYSEEDOROGA 1800
QY 1820 EPRKNFVAPNETKTYFWKQHEMAPTKDEFCCKAWAYFSDVDLEKDVHSGGLGPLLVCHT 1879
Db 1801 EPRKNFVAPNETKTYFWKQHEMAPTKDEFCCKAWAYFSDVDLEKDVHSGGLGPLLVCHT 1860
QY 1880 NTLNFAHGRQVTVQBFALFFITFDETKSWYFTENMERNCRAPCNQMDEPTFKENYRFA 1939
Db 1861 NTLNFAHGRQVTVQBFALFFITFDETKSWYFTENMERNCRAPCNQMDEPTFKENYRFA 1920
QY 1940 INGYMDTLPGLVMAQDORIRWLLSMGSENIHSHFSGHVFTYRKKEEKWALYNLYP 1999
Db 1921 INGYMDTLPGLVMAQDORIRWLLSMGSENIHSHFSGHVFTYRKKEEKWALYNLYP 1980
QY 2000 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKCTPLGMASGHIRDPQITAS 2059
Db 1981 GYFTEVEMLPKAGIWRVECLIGEHLHAGMSTFLVYGNKCTPLGMASGHIRDPQITAS 2040
QY 2060 QYCGOWAPKARLHVSGSINAWSTKEPSWIKVDLLAPMIHGIKTQARQKFSSLYISQ 2119
Db 2041 QYCGOWAPKARLHVSGSINAWSTKEPSWIKVDLLAPMIHGIKTQARQKFSSLYISQ 2100
QY 2120 FIMYSLDGKKWQTRGNSTGLTVFFGNVDSGKKNHFNPIIARVIRLHPHYSIRS 2179
Db 2101 FIMYSLDGKKWQTRGNSTGLTVFFGNVDSGKKNHFNPIIARVIRLHPHYSIRS 2160
QY 2180 TIRBELMGCDLNSCNPGLMESKASDAQITASSYFTNMFTWSPSKARLHLQGRSNWR 2239
Db 2161 TIRBELMGCDLNSCNPGLMESKASDAQITASSYFTNMFTWSPSKARLHLQGRSNWR 2220
QY 2240 PQVNNPKELQVDFORTKWKVTGTTQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2299
Db 2221 PQVNNPKELQVDFORTKWKVTGTTQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
QY 2300 KVFQGNQDSFTPVVNSLDPPLTRYLRIRHPQSVWHQIALRKEVLCGCAQDLY 2351
Db 2281 KVFQGNQDSFTPVVNSLDPPLTRYLRIRHPQSVWHQIALRKEVLCGCAQDLY 2332

RESULT 9

US-10-131-510A-2
; Sequence 2, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Iollar, John S
; TITLE OF INVENTION: Modified Factor VII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26

; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-510A-2

Query Match 98.9%; Score 12282; DB 14; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 20 ATRRYYLCAVELSDYMQSDLGELPVDARPPRPVKPFPFNTSVYKXTLFFVEFTHLNF 79
Db 1 ATRRYYLCAVELSDYMQSDLGELPVDARPPRPVKPFPFNTSVYKXTLFFVEFTHLNF 60
QY 80 IAXRPPPMWGLLGTIIQAEVYDTVITLKNMASHPVSLHVGVSYWKASEGAEDDQTSQ 139
Db 61 IAXRPPPMWGLLGTIIQAEVYDTVITLKNMASHPVSLHVGVSYWKASEGAEDDQTSQ 120
QY 140 REKEDDKVPPGSGTYWQVLKENGPKASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVPPGSGTYWQVLKENGPKASDPLCTYSYLSHVDLVKDLNSGLIGALLVCR 180
QY 200 EGS LAKEKXTQLHFKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 259
Db 181 EGS LAKEKXTQLHFKFILLFAVDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGVNR 240
QY 260 SLPLGLIGHKRSYVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 319
Db 241 SLPLGLIGHKRSYVWHVIGMTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTL 300
QY 320 MDLQGLLSCHISHQHGDGMEAYKVDSCPEPOLIKKNNEEAEDYDDITDSEMDVVRP 379
Db 301 MDLQGLLSCHISHQHGDGMEAYKVDSCPEPOLIKKNNEEAEDYDDITDSEMDVVRP 360
QY 380 DDDNSPFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKSOYLNNGPORIG 439
Db 361 DDDNSPFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKSOYLNNGPORIG 420
QY 440 RYKVKRPFMAYTDETFKTREAIQHESGILGPLLYGEVGDITLLIFKQASRPYNIYPHGI 499
Db 421 RYKVKRPFMAYTDETFKTREAIQHESGILGPLLYGEVGDITLLIFKQASRPYNIYPHGI 480
QY 500 TDVRPLYSRRLPKGVKHLKDPPILPGEIFKYKWTVTVEDGPTKSDPRCLTLYSSFVNME 559
Db 481 TDVRPLYSRRLPKGVKHLKDPPILPGEIFKYKWTVTVEDGPTKSDPRCLTLYSSFVNME 540
QY 560 RDLASGLIGPLLYCYKESVDQGNQIMSDKENVILFSVDENRSMWLTENIQRELPNPAQ 619
Db 541 RDLASGLIGPLLYCYKESVDQGNQIMSDKENVILFSVDENRSMWLTENIQRELPNPAQ 600
QY 620 VQLEDPEFQASNIMHSGVYVDSLOLSVCLHEVAYWVILSIGAQTFLSVFFSGYTFKH 679
Db 601 VQLEDPEFQASNIMHSGVYVDSLOLSVCLHEVAYWVILSIGAQTFLSVFFSGYTFKH 660
QY 680 KMVYEDTILTFPFSGETVFMSENPGLWILGCHNSDFRNQMTALLKVSSCDKXGTGYE 739
Db 661 KMVYEDTILTFPFSGETVFMSENPGLWILGCHNSDFRNQMTALLKVSSCDKXGTGYE 720
QY 740 DSYEDISAYLSKKNNAIEPRFSQNSRHPSTROQFNATTIPENDIEKTDPFWFAHRTMP 799
Db 721 DSYEDISAYLSKKNNAIEPRFSQNSRHPSTROQFNATTIPENDIEKTDPFWFAHRTMP 780
QY 800 KIQNVSSDDLMLLRQSPPTHGLSLSDLOEAKYTFPSDDPSGAIDNNSLSENTHFRPQ 859
Db 781 KIQNVSSDDLMLLRQSPPTHGLSLSDLOEAKYTFPSDDPSGAIDNNSLSENTHFRPQ 840

Qy 860 LHHSGDMVFTDESGLQRLNEKLGTTAATLKKLDQKVSSTSNMLISTIPSDNLAAGTDN 919
Db 841 LHHSGDMVFTDESGLQRLNEKLGTTAATLKKLDQKVSSTSNMLISTIPSDNLAAGTDN 900
Qy 920 TSSIGPPMPVHYDSQDQDTTLPFGKSSPLTBSGGPLSSEENNDKLESGLMKSQBSW 979
Db 901 TSSIGPPMPVHYDSQDQDTTLPFGKSSPLTBSGGPLSSEENNDKLESGLMKSQBSW 960
Qy 980 GKXYSSTESGRLFKGKRAHGALLTKDNALFKVSIISLTKTKTSNNSTNKTTHIDGSL 1039
Db 961 GKXYSSTESGRLFKGKRAHGALLTKDNALFKVSIISLTKTKTSNNSTNKTTHIDGSL 1020
Qy 1040 LIENSPPVQWQILBDETFKVTPLIHDRLMDKXATLRLNHNMSNKTTSKNNEMVQOK 1099
Db 1021 LIENSPPVQWQILBDETFKVTPLIHDRLMDKXATLRLNHNMSNKTTSKNNEMVQOK 1080
Qy 1100 KEGPIPPDAQNPDMSFFKMLFLPESARMIQTHGKNSLNSGGPPKQLVSLGPEKSVEG 1159
Db 1081 KEGPIPPDAQNPDMSFFKMLFLPESARMIQTHGKNSLNSGGPPKQLVSLGPEKSVEG 1140
Qy 1160 QNFISEKNKVVKGFEFTKDYGLKEMVFPSSRNFLTNLDNLHNNTHNQEKKEETEK 1219
Db 1141 QNFISEKNKVVKGFEFTKDYGLKEMVFPSSRNFLTNLDNLHNNTHNQEKKEETEK 1200
Qy 1220 KETLIQENWVLPQIHTVTGCTKNFMKNLFLLSRQNVESYDGAAPVLQDPRSLNDSNR 1279
Db 1201 KETLIQENWVLPQIHTVTGCTKNFMKNLFLLSRQNVESYDGAAPVLQDPRSLNDSNR 1260
Qy 1280 TKKHTAHSKGBEENLGLNCTQIIVKXACTTRISPTNSQONFVQTSKRALKQRL 1339
Db 1261 TKKHTAHSKGBEENLGLNCTQIIVKXACTTRISPTNSQONFVQTSKRALKQRL 1320
Qy 1340 PLEETELEKRIIIVDTSQWKNMHLTSTLTQIDYNEKEKGAITQSPDCLTRSHI 1399
Db 1321 PLEETELEKRIIIVDTSQWKNMHLTSTLTQIDYNEKEKGAITQSPDCLTRSHI 1380
Qy 1400 PQANRSPPIAKVSPFPIRPIYLRVLFDQNSHPLAASVRKDKSGVQESSHFPQGAKK 1459
Db 1381 PQANRSPPIAKVSPFPIRPIYLRVLFDQNSHPLAASVRKDKSGVQESSHFPQGAKK 1440
Qy 1460 NNLSAIITLMTGQREVGSLGTATNSVYKQVNTVLPKPDLPKTSKGVELLPKVEI 1519
Db 1441 NNLSAIITLMTGQREVGSLGTATNSVYKQVNTVLPKPDLPKTSKGVELLPKVEI 1500
Qy 1520 YQKOLPPTETNSGSPCHLDLVEGSLQGTGAIKWEANRPGKVPFLVATESSAKTPSK 1579
Db 1501 YQKOLPPTETNSGSPCHLDLVEGSLQGTGAIKWEANRPGKVPFLVATESSAKTPSK 1560
Qy 1580 LLDPLAWDNHYGTQIPKEBWSQESPEKTAFFKXDTILSLNACSNHAIAINSGQNK 1639
Db 1561 LLDPLAWDNHYGTQIPKEBWSQESPEKTAFFKXDTILSLNACSNHAIAINSGQNK 1620
Qy 1640 EIEVTWAKQGTRELCSQNPVLPKHQREIRTTTQSQOEEDIDYDTISVEMKXDDFIY 1699
Db 1621 EIEVTWAKQGTRELCSQNPVLPKHQREIRTTTQSQOEEDIDYDTISVEMKXDDFIY 1680
Qy 1700 DEDENQSPRSFQKTRHYFIAAVERLWBYGMSSSPHVLNRQAQSGSVQPKVQFQETD 1759
Db 1681 DEDENQSPRSFQKTRHYFIAAVERLWBYGMSSSPHVLNRQAQSGSVQPKVQFQETD 1740
Qy 1760 GSFTOPLYRGELNEHLGLGPIYRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQROGA 1819
Db 1741 GSFTOPLYRGELNEHLGLGPIYRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQROGA 1800
Qy 1820 BRKKNVKNENETKTYFWKQVHHMAITKQBFDCAWAYTSDVLEKDXVHSGLIGPLLVCHT 1879
Db 1801 BRKKNVKNENETKTYFWKQVHHMAITKQBFDCAWAYTSDVLEKDXVHSGLIGPLLVCHT 1860
Qy 1880 NTLNPAHGRQVIVQEFALFTTIFDETKSWYFTENMERNCRAPCNQMEDPTFKENYRFA 1939
Db 1861 NTLNPAHGRQVIVQEFALFTTIFDETKSWYFTENMERNCRAPCNQMEDPTFKENYRFA 1920

Qy 1940 INGVIYMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1999
Db 1921 INGVIYMDTLPGLVMAQDQRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYP 1980
Qy 2000 GVFTVEMLPKAGIWRVECLIGHLHAGMSTLFLVYSNKCQTPLGMAASGHIRDFQTAS 2059
Db 1981 GVFTVEMLPKAGIWRVECLIGHLHAGMSTLFLVYSNKCQTPLGMAASGHIRDFQTAS 2040
Qy 2060 GOYCWAPKLABHYSGSINAWSTKEPFWIKVDLLAPMIHGIKTGACQKPSLSYISQ 2119
Db 2041 GOYCWAPKLABHYSGSINAWSTKEPFWIKVDLLAPMIHGIKTGACQKPSLSYISQ 2100
Qy 2120 FIIMYSLDGKWKQYRGNSGTGLMVFFGNVDSSGKINIENPPIIARIYRLEPHYSIRS 2179
Db 2101 FIIMYSLDGKWKQYRGNSGTGLMVFFGNVDSSGKINIENPPIIARIYRLEPHYSIRS 2160
Qy 2180 TLRMELMGCDLNSCMPLGMBESKAISQAITASSYFTNMFATWSPSKARLHLQGRSNAWR 2239
Db 2161 TLRMELMGCDLNSCMPLGMBESKAISQAITASSYFTNMFATWSPSKARLHLQGRSNAWR 2220
Qy 2240 POVANNPKEMLOVDFOKTKVKTGVTTOGVKSLITSMYVKEPLISSQDGHOWTLFPQNGKV 2299
Db 2221 POVANNPKEMLOVDFOKTKVKTGVTTOGVKSLITSMYVKEPLISSQDGHOWTLFPQNGKV 2280
Qy 2300 KVFQGNQDSFTPVVNSLDPPLITRYLRIHPQSVWQEIARLMEVLGCEAQDIY 2351
Db 2281 KVFQGNQDSFTPVVNSLDPPLITRYLRIHPQSVWQEIARLMEVLGCEAQDIY 2332

RESULT 10

US-10-445-235-2
; Sequence 2, Application: US/10445235
; Publication No. US20040005670A1
; GENERAL INFORMATION:
; APPLICANT: Katherine A. High
; APPLICANT: Rodney M. Canlire
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: CHOP0176
; CURRENT APPLICATION NUMBER: US/10/445,235
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: 60/382,486
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-445-235-2

Query Match 98.9%; Score 12282; DB 15; Length 2332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 20 ATRRYLGAVELSDYMQSDLGELPVDARFPPRPVKPFPNTSVYKTLFVEFTDHLFN 79
Db 1 ATRRYLGAVELSDYMQSDLGELPVDARFPPRPVKPFPNTSVYKTLFVEFTDHLFN 60
Qy 80 IAKRPPPMWGLLGTIOAEVYDVTVTILKNWASHPVSILHAGVSVYWKASGAEYDDQTSQ 139
Db 61 IAKRPPPMWGLLGTIOAEVYDVTVTILKNWASHPVSILHAGVSVYWKASGAEYDDQTSQ 120
Qy 140 REKEDDKVFPGGSHYVQVNLKENGPMASDPLCLTYSLVSHVDLVKDLNSGLIGALLVCR 199
Db 121 REKEDDKVFPGGSHYVQVNLKENGPMASDPLCLTYSLVSHVDLVKDLNSGLIGALLVCR 180
Qy 200 BGS LAKEKTQTLHFIILLFAVDFDECKSWHSETKNSLMDRDAASARAWPKHVTNGYVNR 259
Db 181 BGS LAKEKTQTLHFIILLFAVDFDECKSWHSETKNSLMDRDAASARAWPKHVTNGYVNR 240
Qy 260 SLPLGIGHRKSVTWHVIGMGTTPTEVHSIFLEGHTFLVRNHRQASLEISPTTFTTAQTLL 319

Db 241 SLPGLIGCHRKGVYWEVIGWTTPEVHSIFLBEHTPLVNRHQASLEISPIITFLTAQTL 300
QY 320 MDLGGFLLSCHTSSOHDGWEAYVKVDSCEBPQLIMKNNEAEADYDDLTDSMDVVRP 379
Db 301 MDLGGFLLFCHTSSOHDGWEAYVKVDSCEBPQLIMKNNEAEADYDDLTDSMDVVRP 360
QY 380 DDNSPFSFIQIRSVAKHFKPTWVHYIAAEEEDWDVAPLVLAPDDRSYKSOYLNNPQIRIG 439
Db 361 DDNSPFSFIQIRSVAKHFKPTWVHYIAAEEEDWDVAPLVLAPDDRSYKSOYLNNPQIRIG 420
QY 440 RYKXVRFMAYTDEFTKTEEAIOHSGILGPLLYGEVGTLLIIIPKQASRPYNIYPHGI 499
Db 421 RYKXVRFMAYTDEFTKTEEA-CHESGILGPLLYGEVGTLLIIIPKQASRPYNIYPHGI 480
QY 500 TDVR2LYSR3LPKGVKHLXDFPILPCEIEKYKWTVTVEDGPKSPRCLTRYSSPVMME 559
Db 481 TDVR2LYSR3LPKGVKHLXDFPILPCEIEKYKWTVTVEDGPKSPRCLTRYSSPVMME 540
QY 560 RDLASGLIGPLIICYKESVDQGNQIMSDKXNVILFVSFDENRSWYLTENIORFLPNPAG 619
Db 541 RDLASGLIGPLIICYKESVDQGNQIMSDKXNVILFVSFDENRSWYLTENIORFLPNPAG 600
QY 620 VQLEDPFOASIMHSINGYFDSLSVCLHEVAYWYILSICAQDFLSVFFSGYTFKH 679
Db 601 VQLEDPFOASIMHSINGYFDSLSVCLHEVAYWYILSICAQDFLSVFFSGYTFKH 660
QY 680 KMVYEDTLTLPFSGETVPMSENPCGLWILGCHNSDFRNRGWTALLKVSSCDKNIGDYVE 739
Db 661 KMVYEDTLTLPFSGETVPMSENPCGLWILGCHNSDFRNRGWTALLKVSSCDKNIGDYVE 720
QY 740 DSYEDISAYLLSKNAIEPRFSQNSRHPSTKQKFNATTIPENDIEKTDPMFAHRTMP 799
Db 721 DSYEDISAYLLSKNAIEPRFSQNSRHPSTKQKFNATTIPENDIEKTDPMFAHRTMP 780
QY 800 KIONVSSDMLLKQSPFPHGLSLSDLOEAKYETPSDDPSGCAIDSNNSLSEMTFRPQ 859
Db 781 KIONVSSDMLLKQSPFPHGLSLSDLOEAKYETPSDDPSGCAIDSNNSLSEMTFRPQ 840
QY 860 LHSGDMVFTPSGQLRNEKLGITTAATELKKLDFKVSSTNNLITIPSNLNAAGTUN 919
Db 841 LHSGDMVFTPSGQLRNEKLGITTAATELKKLDFKVSSTNNLITIPSNLNAAGTUN 900
QY 920 TSSLGPPMPHYDOLDTTLFGKXSSPLTESCGPLSSEENNDKLLSEGLMNSQESSW 979
Db 901 TSSLGPPMPHYDOLDTTLFGKXSSPLTESCGPLSSEENNDKLLSEGLMNSQESSW 960
QY 980 GKNVSTESGRIFPKGKRAKGPALLTKDNALFKVSI3LLKTNKTSNNSATNRKTHIDGPSL 1039
Db 961 GKNVSTESGRIFPKGKRAKGPALLTKDNALFKVSI3LLKTNKTSNNSATNRKTHIDGPSL 1020
QY 1040 LIENSPPSVNONTLESDETEPKVTPLIHDRMLMDKNATALRLNHSNKTSSKNMVMVQOK 1099
Db 1021 LIENSPPSVNONTLESDETEPKVTPLIHDRMLMDKNATALRLNHSNKTSSKNMVMVQOK 1080
QY 1100 KEGPIPPDQNDPMSFPKMLFIPESARWIOETHGKNSLNSGQSPKQIVSLGPEKSVEG 1159
Db 1081 KEGPIPPDQNDPMSFPKMLFIPESARWIOETHGKNSLNSGQSPKQIVSLGPEKSVEG 1140
QY 1160 QNFLSEKKNVYKGEFTTQVGLKENVPFSSRNPLFLTNLDNLHENTHNOEKKI0BEIEK 1219
Db 1141 QNFLSEKKNVYKGEFTTQVGLKENVPFSSRNPLFLTNLDNLHENTHNOEKKI0BEIEK 1200
QY 1220 KETLIQENVVLPOIHVTVTGTFKPMKNLFLI3TRQNVGSGYDAGYAPVLQDFRSLNDSTNR 1279
Db 1201 KETLIQENVVLPOIHVTVTGTFKPMKNLFLI3TRQNVGSGYDAGYAPVLQDFRSLNDSTNR 1260
QY 1280 TKXHTAHFKKGBEENLEGLGNQTKQIVKIACTTTRISPNTSQONFVTORSKALKQFRL 1339
Db 1261 TKXHTAHFKKGBEENLEGLGNQTKQIVKIACTTTRISPNTSQONFVTORSKALKQFRL 1320
QY 1340 PLEETELEXRIIVDDTSTOWSKNKHLPSTLTQIDYNEKEKGAITQSPSLDCLFRSHSI 1399
Db 1321 PLEETELEXRIIVDDTSTOWSKNKHLPSTLTQIDYNEKEKGAITQSPSLDCLFRSHSI 1380

QY 1400 POANBSPPIAKVSPFPIRPIYLTVLFOONSSHLPAASYRKKDQSGVQESSHFLQGAKK 1459
Db 1381 POANBSPPIAKVSPFPIRPIYLTVLFOONSSHLPAASYRKKDQSGVQESSHFLQGAKK 1440
QY 1460 NNLSIAIITLWMTGQREVGSIGTSATNSVTYKVKVENTVLPKPDLFKTSGKVELLPKVHI 1519
Db 1441 NNLSIAIITLWMTGQREVGSIGTSATNSVTYKVKVENTVLPKPDLFKTSGKVELLPKVHI 1500
QY 1520 YOKDLFPETNGSPGHLDLVEGSLLOQTEGAIKWNEANRPKVPFLRVAATESAKTFSK 1579
Db 1501 YOKDLFPETNGSPGHLDLVEGSLLOQTEGAIKWNEANRPKVPFLRVAATESAKTFSK 1560
QY 1580 LLDPLAWNHYGTQIPKEBWSOKSPKTAFAKXOTILSNACESHAIIAINEGQNK 1639
Db 1561 LLDPLAWNHYGTQIPKEBWSOKSPKTAFAKXOTILSNACESHAIIAINEGQNK 1620
QY 1640 BIEVTWAKQGRTERLCSQNPVLKXHOBEITRTTILQSDQEBIDYDDTISVEMKKEDFDIY 1699
Db 1621 BIEVTWAKQGRTERLCSQNPVLKXHOBEITRTTILQSDQEBIDYDDTISVEMKKEDFDIY 1680
QY 1700 DBENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVFPQKVVQFQETD 1759
Db 1681 DBENQSPRSQKTRHYFIAAVERLWDYGMSSSPHVLNRQAQSGVFPQKVVQFQETD 1740
QY 1760 GSFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSTYSIIISYEDDOROGA 1819
Db 1741 GSFTQPLVRGELNEHLGLLGPYIRAEVEDNIMVTFRNOASRPYSTYSIIISYEDDOROGA 1800
QY 1820 EPRKNFVKPNETKTYFWKVQHMMAPTKDFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1879
Db 1801 EPRKNFVKPNETKTYFWKVQHMMAPTKDFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHT 1860
QY 1880 NTLNPAHGRQVTVQBPALPFTIFDBTKSWYPTENMERNCRAPCNQMEDPTPKENYRHA 1939
Db 1861 NTLNPAHGRQVTVQBPALPFTIFDBTKSWYPTENMERNCRAPCNQMEDPTPKENYRHA 1920
QY 1940 INGYIMDTLPLGLVMAQDQRIEYLLSMGSENNIHSIHFSGHVFTVRKKEEKYKALYNLYP 1999
Db 1921 INGYIMDTLPLGLVMAQDQRIEYLLSMGSENNIHSIHFSGHVFTVRKKEEKYKALYNLYP 1980
QY 2000 GVFEVTEMLPSKAGIWRVECLJGEBLHAGMSTLFLVYSNKQCTP-LGMAAGHIRDFQITAS 2059
Db 1981 GVFEVTEMLPSKAGIWRVECLJGEBLHAGMSTLFLVYSNKQCTP-LGMAAGHIRDFQITAS 2040
QY 2060 GOYQOWAPKLARLHVSIGSINAWSTKEPFSWIKVDLLAPMIHIGIKTOGARQKFPSSLIYSQ 2119
Db 2041 GOYQOWAPKLARLHVSIGSINAWSTKEPFSWIKVDLLAPMIHIGIKTOGARQKFPSSLIYSQ 2100
QY 2120 FTIMYSLDGKKWQTVRGNSGTGLMVFFGNVDSSGKHNIFNPPIIARVIRLHETHYSIRS 2179
Db 2101 FTIMYSLDGKKWQTVRGNSGTGLMVFFGNVDSSGKHNIFNPPIIARVIRLHETHYSIRS 2160
QY 2180 TLRMBLGMCDLNSCMLPMSKALISDAQITASSYFTNNFATWSPSKARLHLQGSNAWR 2239
Db 2161 TLRMBLGMCDLNSCMLPMSKALISDAQITASSYFTNNFATWSPSKARLHLQGSNAWR 2220
QY 2240 PQVNNPKEWLQVDFOKTKMKTGVTTCQVKSLTTSYVVEFLISSQDGHOWTLFPQNGKV 2299
Db 2221 PQVNNPKEWLQVDFOKTKMKTGVTTCQVKSLTTSYVVEFLISSQDGHOWTLFPQNGKV 2280
QY 2300 KVFQNGQDSFPVNVNSLDPPLLTLYRTHPSQVHQAIALRMEVLGCEAQQDLY 2351
Db 2281 KVFQNGQDSFPVNVNSLDPPLLTLYRTHPSQVHQAIALRMEVLGCEAQQDLY 2332

RESULT 11
US-10-239-498A-2
; Sequence 2, Application US/10239498A
; Publication No. US2004002333A1
; GENERAL INFORMATION:
; APPLICANT: Hauser, Charlotte
; APPLICANT: Horster, Andrea

; APPLICANT: Schroder, Carola
 ; APPLICANT: Lehner, Michael
 ; TITLE OF INVENTION: Production of Recombinant Blood Clotting Factors in
 ; TITLE OF INVENTION: Human Cell Lines
 ; FILE REFERENCE: 80577.0001
 ; CURRENT APPLICATION NUMBER: US/10/239,498A
 ; CURRENT FILING DATE: 2003-07-08
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03220
 ; PRIOR FILING DATE: 2001-03-21
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 2332
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-239-498A-2

Query Match 98.9%; Score 12282; DB 16; Length 2332;

Best Local Similarity 99.8%; Pred. No. 0;
Matches 2327; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy	20	ATRYVYLGAVLSNDYMSDGLGELPVDARPPRPVKSPFFNTSVYKTLFEVETDHLFN	79
Db	1	ATRYVYLGAVLSNDYMSDGLGELPVDARPPRPVKSPFFNTSVYKTLFEVETDHLFN	60
Qy	80	IAKRPPPMGLGPTIOAEVYDVTIVILKNVASHPVSLHAYGVSWKASGAEDDQTSQ	139
Db	61	IAKRPPPMGLGPTIOAEVYDVTIVILKNVASHPVSLHAYGVSWKASGAEDDQTSQ	120
Qy	140	REKEDDKVFPGGSTYVYVQVLKENGPMASDPLCTYSYLSHVDVLKOLNSGLIGALLVCR	199
Db	121	REKEDDKVFPGGSTYVYVQVLKENGPMASDPLCTYSYLSHVDVLKOLNSGLIGALLVCR	180
Qy	200	EGSLAKEKTOIHLKFIILLFAVFDGKSWHSTKNSLMQDRDAASARAKPMKHTVNGYNR	259
Db	181	EGSLAKEKTOIHLKFIILLFAVFDGKSWHSTKNSLMQDRDAASARAKPMKHTVNGYNR	240
Qy	260	SLPLGICHRKSVYVHVGMTTPEVHSIFLEIGHTFLVRNHRQASLEISPTFLTAQTL	319
Db	241	SLPLGICHRKSVYVHVGMTTPEVHSIFLEIGHTFLVRNHRQASLEISPTFLTAQTL	300
Qy	320	MDLQGFLLSCHISFQHDGMZAYVKVDSCEPEQLIMKNNEAEEDYDDDLTDSMDVVRP	379
Db	301	MDLQGFLLSCHISFQHDGMZAYVKVDSCEPEQLIMKNNEAEEDYDDDLTDSMDVVRP	360
Qy	380	DDNSPFIQIRSVAKKHPXTWVHYIAAEEDWDYAPVLAPDDRYSKSYLNKGPQRI	439
Db	361	DDNSPFIQIRSVAKKHPXTWVHYIAAEEDWDYAPVLAPDDRYSKSYLNKGPQRI	420
Qy	440	RKYKVRFMAYTDETFKTRZAIQHSGLILGPLYGEVGDTLIIIFKNQASRPVNIYPHGI	499
Db	421	RKYKVRFMAYTDETFKTRZAIQHSGLILGPLYGEVGDTLIIIFKNQASRPVNIYPHGI	480
Qy	500	TDVRLPYRRLPKGVKHLKDPILPGIFIKYKWTVTVEDGDKSDPRCLTRYYSFVNME	559
Db	481	TDVRLPYRRLPKGVKHLKDPILPGIFIKYKWTVTVEDGDKSDPRCLTRYYSFVNME	540
Qy	560	RDASGLIGPLLI CYKSPVQKQNMISDRKNVILFSVFDNRNRYLTENIQRLPNPAG	619
Db	541	RDASGLIGPLLI CYKSPVQKQNMISDRKNVILFSVFDNRNRYLTENIQRLPNPAG	600
Qy	620	VOLEDPEFQASNMHSINGVYFDSLOLVCHEVAYWYILSIGAQTDFLSVFFSGYTFKH	679
Db	601	VOLEDPEFQASNMHSINGVYFDSLOLVCHEVAYWYILSIGAQTDFLSVFFSGYTFKH	660
Qy	680	KWVYEDTLTLPFPFGSETVMSMENPGLWILGCHNSDFRNKGTALLKVVSCDKNTGDIYE	739
Db	661	KWVYEDTLTLPFPFGSETVMSMENPGLWILGCHNSDFRNKGTALLKVVSCDKNTGDIYE	720
Qy	740	DSYEDISAYLLSKNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTPWFFAHRTMP	799
Db	721	DSYEDISAYLLSKNAIEPRFSQNSRHPSTRQKQFNATTIPENDIEKTPWFFAHRTMP	780

Qy	800	KIQNVSSDILLMLLRQSPTPHGLISLSDLOAKYETFSDDPSGAIDSNNSLSEMTHPRPQ	859
Db	781	KIQNVSSDILLMLLRQSPTPHGLISLSDLOAKYETFSDDPSGAIDSNNSLSEMTHPRPQ	840
Qy	860	LHSGDMVFTPESGIQLRIANEKLGTTAATELKKLDLKFVSSSTNNLISTIPSDNLAAGTDN	919
Db	841	LHSGDMVFTPESGIQLRIANEKLGTTAATELKKLDLKFVSSSTNNLISTIPSDNLAAGTDN	900
Qy	920	TSSLGPPMPVHYDSQDITLFGKSSPLTSGGPLSLSEENDSKLLSEGLMNSQSSW	979
Db	901	TSSLGPPMPVHYDSQDITLFGKSSPLTSGGPLSLSEENDSKLLSEGLMNSQSSW	960
Qy	980	GKNTSSTESGRLFKGQAAGPALLITKONALPKVSIILKTKNTKSNNAATNKTTHIDGSL	1039
Db	961	GKNTSSTESGRLFKGQAAGPALLITKONALPKVSIILKTKNTKSNNAATNKTTHIDGSL	1020
Qy	1040	LIENSPSVWQNILSDTEFKKVTPLIHDRMLMDKNATALRLNHSNKTTSKMEMVQOK	1099
Db	1021	LIENSPSVWQNILSDTEFKKVTPLIHDRMLMDKNATALRLNHSNKTTSKMEMVQOK	1080
Qy	1100	KEGPIPPDAQNPDMSFFKMLFLPESARWIQETHGKNSLNSCQGPSKQVLVSGPEKSVEG	1159
Db	1081	KEGPIPPDAQNPDMSFFKMLFLPESARWIQETHGKNSLNSCQGPSKQVLVSGPEKSVEG	1140
Qy	1160	QNFISEXNKVVVGKERTKDVGLKEMVFPSSRNFLTNLMDLNHNTHNQSKTKQETIEK	1219
Db	1141	QNFISEXNKVVVGKERTKDVGLKEMVFPSSRNFLTNLMDLNHNTHNQSKTKQETIEK	1200
Qy	1220	KETLIQENVVLPOIHTVTGTQNFKNFLFLSTRQNVESYDGAYAPVLQDPRSLNDSTNR	1279
Db	1201	KETLIQENVVLPOIHTVTGTQNFKNFLFLSTRQNVESYDGAYAPVLQDPRSLNDSTNR	1260
Qy	1280	TKKHTAFHSKGBEENLEGIGNOTKOIYEKACVTRISPNTSQONFVQSKRALKQFRL	1339
Db	1261	TKKHTAFHSKGBEENLEGIGNOTKOIYEKACVTRISPNTSQONFVQSKRALKQFRL	1320
Qy	1340	PLEBTELEKRIIVDDTSTQMSKMKHLTPSLTQIDYNEKEKGAITOSPISDCLTRSHSI	1399
Db	1321	PLEBTELEKRIIVDDTSTQMSKMKHLTPSLTQIDYNEKEKGAITOSPISDCLTRSHSI	1380
Qy	1400	POANRSLPITAKVSSPFSIRPIYITRVLFQDNSSHLPAASVYRKDKSGVQSSSHFLQAKK	1459
Db	1381	POANRSLPITAKVSSPFSIRPIYITRVLFQDNSSHLPAASVYRKDKSGVQSSSHFLQAKK	1440
Qy	1460	NNLSAILTLTLEMTQDQREVSLGTSATNSVYKVENTVILPKDLPKTSKGVLLPKVHI	1519
Db	1441	NNLSAILTLTLEMTQDQREVSLGTSATNSVYKVENTVILPKDLPKTSKGVLLPKVHI	1500
Qy	1520	YQKDLPTETSNSSPGHLDLVEGSLLOGTEGAIKWNEANRPGKVPFLRVATESSAKTPSK	1579
Db	1501	YQKDLPTETSNSSPGHLDLVEGSLLOGTEGAIKWNEANRPGKVPFLRVATESSAKTPSK	1560
Qy	1580	LLDPLAWNHYGTQIPKEEWKSOEKSPKTAFFKKDTILSLNACESHAAANNEGONKP	1639
Db	1561	LLDPLAWNHYGTQIPKEEWKSOEKSPKTAFFKKDTILSLNACESHAAANNEGONKP	1620
Qy	1640	EIEVTWAKQRTERLCSONPPVLKHOREITRTTLTQSDQDEIDYDDTISVEMKKEDEIY	1699
Db	1621	EIEVTWAKQRTERLCSONPPVLKHOREITRTTLTQSDQDEIDYDDTISVEMKKEDEIY	1680
Qy	1700	DEBENQSPRSFQKTRHYFTAAVERLWDYGMSSPHVLRNRAQSGSVQPKVVFQFTD	1759
Db	1681	DEBENQSPRSFQKTRHYFTAAVERLWDYGMSSPHVLRNRAQSGSVQPKVVFQFTD	1740
Qy	1760	GSTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFNQASRPYSFYSSLSIYSEDORQGA	1819
Db	1741	GSTQPLRYGELNEHLGLGPIYRAEVEDNIMVTFNQASRPYSFYSSLSIYSEDORQGA	1800
Qy	1820	EPKNTFKVNETKTYFWKVOHMAKPKDEPDKAWAYFSDVLEKDVHSLIGPLVCHT	1879
Db	1801	EPKNTFKVNETKTYFWKVOHMAKPKDEPDKAWAYFSDVLEKDVHSLIGPLVCHT	1860
Qy	1880	NTLNPAHGRQVTOEFALPFTTIFDETKSWYFTENMBENCRAPCNIQMEDPTFKENYRFAH	1939

1861 NTLNPAHQRTVQBFALFFIFDETKGYSFTENMERNCRAPNTQMEDPTPKENYRTHA 1920
1940 INGVIMDTLPGVMAQDQRIKWLMSGNSNENIHSIHSGVFTVRKKEEYKALYNLYP 1999
1921 INGVIMDTLPGVMAQDQRIKWLMSGNSNENIHSIHSGVFTVRKKEEYKALYNLYP 1980
2000 GYFTEVEMLPKAGIWRVECLTGEHLHAGMSTLFLVYGNKQOTPLGMASGHIRDQITAS 2059
1981 GYFTEVEMLPKAGIWRVECLTGEHLHAGMSTLFLVYGNKQOTPLGMASGHIRDQITAS 2040
2060 QOYGOWAPKARLHYSGSINAWSTKEPFSWIKVULLAPMIHGIKTOGARQKPFSLYISQ 2119
2041 QOYGOWAPKARLHYSGSINAWSTKEPFSWIKVULLAPMIHGIKTOGARQKPFSLYISQ 2100
2120 FIIMYSLDGKQWOTVRGNSGTGLMVFVGNVDSSGKHNIEPPIIARVIRLHPHTHYSIRS 2179
2101 FIIMYSLDGKQWOTVRGNSGTGLMVFVGNVDSSGKHNIEPPIIARVIRLHPHTHYSIRS 2160
2180 TLRMELMGCDLNSCSPMLGMSKALSDAQITASSYFTNWFATWSPSKARLHLQGRSNARW 2239
2161 TLRMELMGCDLNSCSPMLGMSKALSDAQITASSYFTNWFATWSPSKARLHLQGRSNARW 2220
2240 POVANPKWLQVDFQTKMKVGTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2299
2221 POVANPKWLQVDFQTKMKVGTGVTQGVKSLTSMYVKEFLISSQDGHQWTLFPQNGKV 2280
2300 KVFQGNQDSFTPVNSLDPPLTLRLYRHPQSWHVOIALRMEVLGCEAODLY 2351
2281 KVFQGNQDSFTPVNSLDPPLTLRLYRHPQSWHVOIALRMEVLGCEAODLY 2332

RESULT 12

US-10-187-319-6
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/187,319
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/523,656
; FILING DATE: 2000-03-10
; APPLICATION NUMBER: US 09/037,601
; FILING DATE: 1998-03-10
; APPLICATION NUMBER: WO PCT/US97/11155
; FILING DATE: 1997-06-26
; APPLICATION NUMBER: US 08/670,707
; FILING DATE: 1996-06-26
; ATTORNEY/AGENT INFORMATION:
; NAME: Greenlee, Lorraine L.
; REGISTRATION NUMBER: 27,894
; REFERENCE/DOCKET NUMBER: 75-95X
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; TITLE: Sequence of the Murine Factor VIII cDNA
; JOURNAL: Genomics
; VOLUME: 16

PAGES: 374-379
DATE: 1993
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-187-319-6
Query Match 71.0%; Score 8820; DB 14; Length 2319;
Best Local Similarity 72.4%; Pred. No. 0;
Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;
QY 1 MQIELSTCFCLLRPCFSATRRYYLGAVELSWPMQSD-IGELPVDARPPRPVKSPFF 59
Db 1 MQIALFAFCFLSLFNFCSSAIRYYLGAVELSWNYIQSDLLSVLHSDSRFLPRMSTSPFF 60
QY 60 NTSVYVYKKTAFVBTDLHLNIAKRPWGMGLLGTIOAEVYDVVITLKNMASHPVSLHA 119
Db 61 NTSIMYKKTAFVEYKQDLFNIAKRPWGMGLLGTITWTEVHDTVVITLKNMASHPVSLHA 120
QY 120 VGSYWKASGAEYDDDTQSREKEDCKVFPQGSHTYVMQVLKENGPMASDPLCLTYSL 179
Db 121 VGSYWKASGAEYDDDTQSREKEDCKVFPQGSHTYVMQVLKENGPMASDPLCLTYSL 180
QY 180 HVDLVKDLNSGLIGALLVCREGSLAKTQPLHFKFILLFAVDEGKSWHSTKNSLMQDR 239
Db 181 HVDLVKDLNSGLIGALLVCKEGSLKERTQMLYQVLLFAVDFDEGKSWHSTNDSTQSM 240
QY 240 DAASARAPWKHTYNGVYVNSLPLGLIGCHRKSVMYHVGTTTPVHSIPLFEGHTELVRN 299
Db 241 DSASARAPWKHTYNGVYVNSLPLGLIGCHRKSVMYHVGTTTPVHSIPLFEGHTELVRN 300
QY 300 HRQASLEISPTITLTAQTLMLDGLQFLLSCHSHSHQHDGMEAYVYKVDSCPEPOLIMK-N 358
Db 301 HRQASLEISPTITLTAQTLMLDGLQFLLSCHSHSHQHDGMEAYVYKVDSCPEPOLIMK-N 360
QY 359 NEBAEDDDDLTDSEMDVVRPDDNSPSFTQIRVAKKHPTWVHYTAABEDWDYAPLV 418
Db 361 NEEMEDYDDDLT-SEMDMFTLDYDSSP-FQIRSVAKKYPKTHIYISAEEDWDYAPSV 418
QY 419 LAPDRSVKSOYLANGPQIRGRKYKVRFMAYTDETFKTRAIQHESGILGPLYGEVD 478
Db 419 PTSNGSYKSOYLSNGPHRIGRKYKVRFIAYTDETFKTRAIQHESGILGPLYGEVD 478
QY 479 TLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPGVKHLDKDFPILPGEIFKYKWTIVED 538
Db 479 TLLIIFKNQASRPYNIYPHGITDVSPLHARRLPGIKHVKDLPHPGEIPKYKWTIVED 538
QY 539 GPTKSDPRCLTRYYSFVNMERDLASGLIGPLLIICYESVDQKNQIMSKRNVLISVF 598
Db 539 GPTKSDPRCLTRYYSFVNMERDLASGLIGPLLIICYESVDQKNQIMSKRNVLISVF 598
QY 599 DENRSWYLTENIQRFNLPAGVQLEDPEFOASNTMHSINGVYVFDLSQVLCHEVAYWYI 658
Db 599 DENQSWITENMQRFNLPAAKTQPDQFGQASNMHSINGVYVFDLSQVLCHEVAYWYI 658
QY 659 LSIQAQTDVLSVFPFGYTFKHMYEDTLTLFPSPGSETVFMENPGLWILGCNSDFRN 718
Db 659 LSVGAQTDVLSVFPFGYTFKHMYEDTLTLFPSPGSETVFMENPGLWILGCNSDFRN 718
QY 719 RGMNTALLKVSSCDKNTGDYEDSDYEDISAVLLSKNNAIEPESFQNSRHSRSTROKQFAT 778
Db 719 RGMNTALLKVSSCDKNTGDYEDSDYEDISAVLLSKNNAIEPESFQNSRHSRSTROKQFAT 778
QY 779 TIPENDIEKTDPMFAHATPMFKIQNVSDDLMLLROS-PTEHGLSLSDLOEAKYETFS 837
Db 779 TIPKNDMEKIEPQEEIEAEMLKQSVSVSDMLMLLQSHHTPHGLFLSDQOEALYEIHD 838
QY 838 DPSPGAIIDSNNLSJEMTHFRPQLHSGDMVFTPESGQLRLNEXLGTATAELKKLDFKV 897
Db 839 DHSNADSDNCGPSKVTQLRPESHHSEKIVFTPPQGLQLRSNKSLETTIEVKWKKGLQV 898
QY 898 SSTSNLLI-STIPSDNLAAGTDNTSSIGPSPMPVHYDSQDITLPGKSSPLTEBSGGLS 956
Db 899 SSLEFSLMTTILSDNLKATFEKTDSSGFFDMPYHSSSKSLSTTAFGKKAYSLVGSHPFN 959

Qy 957 LSEBNDKLLSEGMNSQESSWGNVSTSTSCPLFKCKRAFGGALLTKONALFKVSTSL 1016
Db 959 ASEENSDNLDSTLMYSQESPLRDLNLSIENDLEKREPHGIALTKONTLFDKNVSL 1018
Qy 1017 LKTKNTSNSTNKTTHDGSLSLSTSPSVWQN--ILESDFEKKVPLIHDRLMLDKNA 1075
Db 1019 MKTKNTYHSTNEKLHTSEST--SIENSTDLQAILKVNSEIOEVTALIHGDTLLGKNS 1077
Qy 1076 TALRLNMSNTTSGSKNMWVQCKEGBIPDDAQNDKVPFKMLPLPESARWIOETHGKN 1135
Db 1078 TYLSLHMLNLTSTGKNDI--ZHRKQEDPIQDEENTIMPFKMLFLSESSWPFKTKNGN 1137
Qy 1136 SLNSQGPSKCLVSLGPEKSVGEQNEFSEKKNVVGKGETKOVGLKEMVPPSSRLFL 1195
Db 1138 SLNSEQEHSQOLVYLMFKVKNQSFUSEKKNVVEQGGTKNIGLKDMAPPHNMSIFL 1197
Qy 1196 TLNOLNHNTHNQFKKIQEBIEKETHIQENVLFPQIHTVTGTNFMKNLFLASTQNV 1255
Db 1198 TLLSNVHENGHNQEKIQEBIE--KEALIEKVLPQVHEATGSKNFKLDILIGTRQNI 1256
Qy 1256 EGSYDGAAPVLQDFRSLNDSTNRKTKTAHPSK--KGEENLEGLNQTKQIVKCYACT 1313
Db 1257 --SLYEVHVPVLQNTITSNNSTNTVQIEHMFHFKRKDKETNSEGLVNKTRFEMVNY--- 1311
Qy 1314 TRISNTSQNFVOTORSKRAKQFRLPLEETELEKRI--VDDTSTOWSKNMKHLTPSTLTQ 1373
Db 1312 -----PSQKNTTORSKRAKQFRL-----STOWLKTINCSOCIKQ 1349
Qy 1374 IDYNEKEGALITQSLDCLTRSHSIPQANRSPPIAKVSPSPSIRPIYLRVLFDQNS 1433
Db 1350 IDHSEKMKFITKSLSDS--SVIKSTTQNTSSDSHVKTSAFP---PIDLKRSPQNKFS 1405
Qy 1434 HLPAAZY-----RKDSGVQBSHFLOGAKONLSLAITLEMTGQREVSGLSGTSATNSV 1489
Db 1406 HVQASSYTYDFKTSRQIESNNFLKTKINNPSTAILPWNFIDQKFTSPGKNTNSV 1465
Qy 1490 TYKKVENTVLKPDLPKTSKGVKELLPKVHIYKOLPFTETNSGSPGHLDELVEGSLQOTE 1549
Db 1466 TYKKENTIFLKPTEPESGKIELLPQVSIQEBILPTETSHGSPGHLNLMKEVLFQIK 1525
Qy 1550 GAIKNEANRFGKVPFLRVATESSAKTPSKLLDPLAWNHYQTQPKBEWKSQKSPKT 1609
Db 1526 GPTKWKAKRHE--S1KGKTSSKNTSKLNLHAWDYVHAAQPKDMWKSKEKSPRII 1583
Qy 1610 AFKKDITLSLNACESNEAIAINECONKPEVETWAKQGRTERLCSQNPVYLKBSQEI 1669
Db 1584 STKQEDTILSRPHGNSHISGA--NKKQNPQRETTWVQQTQRTCSQPPVVKRQREL 1642
Qy 1670 TRTT--QSQDEIDYDITISVENMKKEDFDIYDBDENQSPRSFQKTRHYPIAAVERLWDY 1729
Db 1643 --SAFQSQEANTDYDAITIE--TIEDFDIYSEDIKQGPRESFQKTRHYPIAAVERLWDY 1699
Qy 1730 MSSSPHVLNRAQSSVPOFKKVVQBEFTDGSFTOPLRGELNEHLGLLGPVIRAEVDN 1789
Db 1700 MST--HVLNRYQSDNVPOFKKVVQBEFTDGSFQPLRGELNEHLGLLGPVIRAEVDN 1758
Qy 1790 INVTFRNOASRYSYFSSLSIYSEEDORQGAERKQFVKENETKTYFWKVOHMAPTKDF 1849
Db 1759 INVTFRNOASRYSYFSSLSIYSEEDORQGAERKQFVKENETKTYFWKVOHMAPTKDF 1817
Qy 1850 DCKAWAYFSDVLEKDVHSGSL--GPLLAVCHTNTLNPAHGRQVTVQEFALFTTFIDETKSY 1909
Db 1818 DCKAWAYFSDVLEKDVHSGSL--GPLLAVCHTNTLNPAHGRQVTVQEFALFTTFIDETKSY 1877
Qy 1910 FTENMERNOCAPCNIOMEDPTKENTRFAALNGYINDTLPGLVMAQDQIRVYLLSMGN 1969
Db 1878 FTENMERNOCAPCNIOMEDPTKENTRFAALNGYINDTLPGLVMAQDQIRVYLLSMGN 1937
Qy 1970 ENIESHFSGHVTFTVKKSEYKMALYNLPGVFTVEMLPKSGAGIWRVCLLGEHLHAGM 2029
Db 1938 ENIQSIHSGHVTFTVKKSEYKMAVYNLPGVFTVEMLPKSGAGIWRVCLLGEHLHAGM 1997

Qy 2030 STLFLVYSNKOTPLMGASHIRDFQITASQYQGWAPKLARLHYSGSINAWSTKEPFSW 2089
Db 1998 STLFLVYSKCOQPLGNASOSIRDFQITASGHYQGWAPKLARLHYSGSINAWSTKEPFSW 2057
Qy 2090 IKVDLLAPMIHGIKTQGARQKFSLSYISOFIIMYSIDGKKWQTYRGNSTGTLVVFQNV 2149
Db 2058 IKVDLLAPMIHGIKTQGARQKFSLSYISOFIIMYSIDGKKWQTYRGNSTGTLVVFQNV 2117
Qy 2150 DSSGIKENIENPPIIARYIRLHPHYRSTIRLAMELGCNLSMPLGMEKSAISDAQI 2209
Db 2118 DSSGIKENIENPPIIARYIRLHPHYRSTIRLAMELGCNLSMPLGMEKSAISDAQI 2177
Qy 2210 TASSYFTNMFATWSPSKARLHLOGRSNARPNQVNNPKNWQVDFQKTMKVTVTQGVKS 2269
Db 2178 TASSYFTNMFATWSPSKARLHLOGRSNARPNQVNNPKNWQVDFQKTMKVTVTQGVKS 2237
Qy 2270 LLTSMYKBEPLISSQGHQWTLFPQNGKVKVFOGQNDSTFPVNSLDPPLITRYLRHP 2329
Db 2238 LFTSMYKBEPLISSQGHQWTLFPQNGKVKVFOGQNDSTFPVNSLDPPLITRYLRHP 2297
Qy 2330 QSWVHOJALRMEVLGCEAQQDLY 2351
Db 2298 QWBEHQALRDEILGCEAQQY 2319
RESULT 13
US-10-131-510A-6
; Sequence 6, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCI/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCI/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 2319
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-131-510A-6
Query Match 71.0%; Score 8820; DB 14; Length 2319;
Best Local Similarity 72.4%; Pred. No. 0;
Matches 1711; Conservative 232; Mismatches 365; Indels 54; Gaps 22;
Qy 1 MQEILSTCFFLCJLAFPCFSATRRYYLGAVELSWDMQSD--IGELPVDARPPRPVKSPFP 59
Db 1 MQIALFACFFLSLNFCSAIRRYLGAVELSWNYIQSDLLSVLHSDSRFLPRMSTSPFP 60
Qy 60 NTSVYVKKTLFVEPFDDELFNIAKPPPMWGLLGPITQAEVYDVTVITLKNMASHVSLHA 119
Db 61 NTSIMYKTKTVFYEYKQDLFNIAKPPPMWGLLGPITQAEVYDVTVITLKNMASHVSLHA 120
Qy 120 VGVSYWKASGAEYDDQSQSREKDDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLS 179
Db 121 VGVSYWKASGAEYDDQSQSREKDDKVPFGSGSHYVWQVLKENGPMASDPLCLTYSYLS 180

DB	2298 QIWEHQIALRLEILGCCAQOQY 2319	DB	241 DPAPARQAPMHTVGVYVNRSLPGLIGCHKKSYVHVIWGMTSPVHVSIFLEGHITFLVRH 300
RESULT 14		QY	300 HROASLEISPIITFLTAQTLMDLQGFLLSCHISHOHGMEAVKVVDSPEEPQOLIMQNN 359
US-10-187-319-37		DB	301 HROASLEISPIITFLTAQTLMDLQGFLLCHISSEHGGMEAVRVESCAEESPQLRRKAD 360
Sequence 37, Application US/10187319		QY	360 EEAEDYDDDLTDSMDVVRPDDNSPSFIOIRSVAKKHPKTWHYIAAEEEDWDYAPLV 419
Publication No. US20030068785A1		DB	361 EE-EDYDNLVDSMDVRLDGDVSPFIOIRSVAKKHPKTWHYIAAEEEDWDYAPVP 419
GENERAL INFORMATION:		QY	420 APDDRSYKSOYLNNGPQIRGKYKVFMAYTDETKTREAIQHEGILGLPLLYGEVGT 479
APPLICANT: Lollar, John S.		DB	420 SPDSRSYKSLYNSGPQIRGKYKARFVAYTDTPTKIRKAIPEYSGILLGLPLLYGEVGT 479
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII		QY	480 LLIIFKQASRPYNIYPHGITDVPYLSRLPKGVHLLKDFPILPGEIKPKYKWTWVEOG 539
NUMBER OF SEQUENCES: 40		DB	480 LLIIFKQASRPYNIYPHGITDVSALHPGLLKGWHLKDMPLPGETFKYKWTWVEOG 539
CORRESPONDENCE ADDRESS:		QY	540 PTKSDPRCLTRYYSFVNMRDLASGLIGPLLICYKESVDQKGNQINMSDKRNVLFSVFD 599
ADDRESSES: Greenlee, Winner and Sullivan, P.C.		DB	540 PTKSDPRCLTRYYSINLEKDLASGLIGPLLICYKESVDQGNQMSDKRNVLFSVFD 599
STREET: 5370 Manhattan Circle Suite 201		QY	600 ENRSWILTENTQRFPLNPAGVQLEDPEQASNMHISINGVVFSLQSLCVLHEVAWYIL 659
CITY: Boulder		DB	600 ENQSWYLAENTQRFPLNPAGVQLEDPEQASNMHISINGVVFSLQSLCVLHEVAWYIL 659
STATE: Colorado		QY	660 SIGAOTDFLSVFFSSYTFKHMVYEDTLTLFPFSGETVFMSPENPGLWILGCHNSDFRNR 719
COUNTRY: USA		DB	660 SVGAOTDFLSVFFSSYTFKHMVYEDTLTLFPFSGETVFMSPENPGLWILGCHNSDLNR 719
ZIP: 80303		QY	720 GMTALLKVSSCDKNTGYDESDYEDISAYLLSKNAIEPRSPQNSRHPSTROKQNAIT 779
COMPUTER READABLE FORM:		DB	720 GMTALLKVSSCDKNTGYDYNTYEDIPGFLLSGKNVIEPRSPQNSRHPSTROKQNAIT 779
VEIDIUM TYPE: Floppy disk		QY	780 IPENDIEXTDPMFAHRTMPKIQNVSSDULMLLRQSPTEHGLSLDQAKYETETSDDP 839
COMPUTER: IBM PC compatible		DB	780 SPEDDVE-LDPQSGERTQALBELSPVSDGSMGLQNPAPHGSSSSDQBARNE--ADDY 836
OPERATING SYSTEM: PC-DOS/MS-DOS		QY	840 SPGAIDSNNSISEMTHFRPQLHSGDMVFTPEESGLQRLNEKLGTTAATLKKLDKPVSS 899
SOFTWARE: Patentin Release #1.0, Version #1.30		DB	837 LFGARERTAPSAAARLRLPHHSAERVLTEP-----EK-----ELKKLDKGMSS 882
CURRENT APPLICATION DATA:		QY	900 TSNNLIS--TIPSDNLAAGTNTSILGPPSPMPHYDSDLODTTLFGKXSSPLTBSGGPLSL 957
APPLICATION NUMBER: US/10/187,319		DB	883 SSDLLKTSPTIPSDTLSEAETRTHSLGPHQVNFRRSOLGAIVLGNKNSHIFGAGVPLGS 942
FILING DATE: 27-Aug-2002		QY	958 SEENNDKSLLESGLMNSQESSGNKXVSTESGRIFKGRAGHPALLTKDNALFKVSLSL 1017
CLASSIFICATION: <Unknown>		DB	943 THEDH-----ESSIGENVSPVESDGFEXEKAHGPASLTCTDDVLFKWNISLV 999
PRIOR APPLICATION DATA:		QY	1018 KINKTNNNSAENRKHIDGPILLTENSFVWQNTLESDFTEFKVTPILHDBMLDKNATA 1077
APPLICATION NUMBER: US 09/523,656		DB	990 KTKARVVLKTNRKTHIDDAALLTENRASA-----TFMDKNTTA 1028
FILING DATE: 2000-03-10		QY	1078 LNLNHSNKTTSKKNWVQCKE3TIPDQAQNDMSFFKMLFLPESARWIOETHGKNSL 1137
APPLICATION NUMBER: US 09/037,601		DB	1029 SGLNHSVN-----WIKGPLGNPL 1047
FILING DATE: 1998-03-10		QY	1138 NSGQGPSKQLVSLGPEKSVGEONFLSEKKNVYVKGFTKQVGLKMWPPSPGRNPLFTN 1197
APPLICATION NUMBER: WO PCT/US97/1-155		DB	1048 SSERGSPPELLTSSGSGKSVKSGSQGQIRVAVEEBELSKG---KEMMLENSELFTLN 1104
FILING DATE: 1997-06-26		QY	1198 LONLHNTHNCKEAKIOEEIEKKTLEQENVVLQIHTVTGTKNFMKNLFLSTTRQNVG 1257
APPLICATION NUMBER: US 08/670,707		DB	1105 SADVQNDTHSCKKREEMERREKLVQKVLQPVYATGATKPKFLNRIHQSTPEPSVEG 1164
FILING DATE: 1998-06-26		QY	1258 SYDGAAPVLQDPRSLNDSTNRKTKHTAHPKSGEENLEGLNQTKQIVKYACTTRIS 1317
ATTORNEY/AGENT INFORMATION:		DB	1165 FDGGSHPVPQDSRLNDSAREAEATHIAFTSAIREAPLEAPGNET----- 1210
NAME: Greenlee, Lorraine L.		QY	1318 PNTSQNFVTCSSKRALQKQFRILPLEETELEKRIIVDDTSTOWSKMKHLTPSLTQIDYN 1377
REGISTRATION NUMBER: 27,894			
REFERENCE/DOCKET NUMBER: 75-95K			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: 303/499-8080			
TELEFAX: 303/499-8089			
INFORMATION FOR SEQ ID NO: 37:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 2133 amino acids			
TYPE: amino acid			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
SEQUENCE DESCRIPTION: SEQ ID NO: 37:			
US-10-187-319-37			
Query Match	64.9%; Score 8057; DB 14; Length 2133;		
Best Local Similarity	66.9%; Pred. No. 0;		
Matches	1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;		
QY	1 MQEILSTCFLLCRFCFATRYVILGAVELSDWYQSD-LGELPVDARPPRPVKSPFP 59	QY	1029 SGLNHSVN-----WIKGPLGNPL 1047
DB	1 MQEILSTCFVLCPLPGFSAIRYVILGAVELSDWYRQSELRLHLDVTRFPATAPGALPL 60	QY	1138 NSGQGPSKQLVSLGPEKSVGEONFLSEKKNVYVKGFTKQVGLKMWPPSPGRNPLFTN 1197
QY	60 NTSVYVKTLFVFTDLFNIAKPRPVMGLGPTIQAEVYTVVITLKNMASHPVSLHA 119	DB	1048 SSERGSPPELLTSSGSGKSVKSGSQGQIRVAVEEBELSKG---KEMMLENSELFTLN 1104
DB	61 GPSVLVYKTVFVFTQPLFSVAPRPVKKVGLGPTIQAEVYTVVITLKNMASHPVSLHA 120	QY	1198 LONLHNTHNCKEAKIOEEIEKKTLEQENVVLQIHTVTGTKNFMKNLFLSTTRQNVG 1257
QY	120 VGVSYWKAEGAYDDQTSOREKEDKVPFGSGSHYVVOVLKENGPMASDPLCLTYSYLS 179	DB	1105 SADVQNDTHSCKKREEMERREKLVQKVLQPVYATGATKPKFLNRIHQSTPEPSVEG 1164
DB	121 VGVSWKSSGAEYDHTSOREKEDKVLFGKSQTVVVOVLKENGFTASDPFCLTYSYLS 180	QY	1258 SYDGAAPVLQDPRSLNDSTNRKTKHTAHPKSGEENLEGLNQTKQIVKYACTTRIS 1317
QY	180 HVDLVKDLNGLGALLVCHREGSLAKETQTLHKFTLLFAVDEGKSWHSETKNSLMQDR 239	DB	1165 FDGGSHPVPQDSRLNDSAREAEATHIAFTSAIREAPLEAPGNET----- 1210
DB	181 HVDLVKDLNGLGALLVCHREGSLTRERTQNLHFEVLLFAVDEGKSW		

```
Db 1211 -GPGPRSAVPRVKOSLKOIRLPLLEBIKPERGVNLNATSTRWS----- 1252
QY 1378 EKEKGAITQSPLSDCLTRSHSIPQANRSLPIAKVSSPFSIRPIYLTRVLFDQNSHLLPA 1437
Db 1253 -----BSSPLQAKRNLSLPLTLEMAGGKGKISALGKSAAGPLASGLJKA 1301
QY 1438 ASYRKDCSGVSSSHFLQAKKNLSLAILTEMTGDQREVSLGFSATNSVYTKKVENT 1497
Db 1253 -----BSSPLQAKRNLSLPLTLEMAGGKGKISALGKSAAGPLASGLJKA 1301
QY 1498 VLPKPDLPKTSKVELLPKHVHYQKDLFTETNSGSPGHLDLVBGSLGCTGALKWNEA 1557
Db 1302 VLSAGLSASGKAESLPKVRHREDLLPQKTSNVSCAHDGLQETFLQKTLGPNVLNKV 1361
QY 1558 NRPCKVPFLRVATSSAKTPSKLLDPLAWDNHVGTOIPKEEMKSPKSPKTAFFKKDPI 1617
Db 1362 NRP-----RTPSKULGP-----PMPK-EWESLEKSPKSTALRKDII 1398
QY 1618 -ISLACESNHAIAINEQNKEPEIEVTWAKQGRTERLCSQNPPLVKRQREITRTLOS 1676
Db 1399 SLPLDRHESNHSIAAKNEQAETQREAAWTQGGPGRLCAPKPPVLRREQRDISLPTQ 1458
QY 1677 DQESDYDDTISVEMKEKEDDIYDENQSPRSFQKTRHYFIAAVERLWDYGMSSSPHV 1736
Db 1459 EDKMDYDDIFSTETKGEDFDIYDENQDPRSFKQRTHYFIAAVEQLWDYGMSESPZA 1518
QY 1737 LNRAGSGSVPOPKVVOEFTDGSFTQPLRGELNEHLGLGPYIRAEVDNIMVTFN 1796
Db 1519 LNRAGNGEVPRFKVVPREFADGSGFTQSYRGELNKHLLGLGPYIRAEVDNIMVTFN 1578
QY 1797 QASRPYSYSSLSIYSEEDQOQAEPKRNFKVKNETXTYFWKVOHMAPTKDFDCWAY 1856
Db 1579 QASRPYSYSSLSIYSPDDQEQGAEPKRNFKVKNETXTYFWKVOHMAPTKDFDCWAY 1638
QY 1857 FSDVLEKDVHSLGPIPLVCHTNTNPAHGRVTVQEFALPTIIFDETKSYFTENMR 1916
Db 1639 FSDVLEKDVHSLGPIPLVCHTNTNPAHGRVTVQEFALPTIIFDETKSYFTENMR 1698
QY 1917 NCRAPCNIMQEDPTPKENYRPHAINGYIMDTLPGLVMAQDQRIRWLLSGNSNENIHS 1976
Db 1699 NCRAPCHLQVEDPTLXENYRPHAINGYIMDTLPGLVMAQDQRIRWLLSGNSNENIHS 1758
QY 1977 FSGHVTVRKKEBYKALYNLYPGVETVEMLPKAGIWRVECLIGEHLHAGNSTLFLVY 2036
Db 1759 FSGHVTVRKKEBYKAVYNLYPGVETVEMLPKAGIWRVECLIGEHLHAGNSTLFLVY 1818
QY 2037 SNKQPLGMAHGIRDFOITASGOYGOWAPKLABLHYSGINAWSTKPPSWIKVDLLA 2096
Db 1819 SNKQPLGMAHGIRDFOITASGOYGOWAPKLABLHYSGINAWSTKPPSWIKVDLLA 1878
QY 2097 PMLIHGIKTQGARQKTSLSIYSQFIIMYSLDGKKWQTYRGNSTGTLWVFGNVDSGIXH 2156
Db 1879 PMLIHGIKTQGARQKTSLSIYSQFIIMYSLDGNWQSYRGNSTGTLWVFGNVDSGIXH 1938
QY 2157 NFNPPIIARYIRLHTHTHSIRSLTRELMLGCDLNSCMPLGMSKAISDAQITASSYFT 2216
Db 1939 NFNPPIIARYIRLHTHTHSIRSLTRELMLGCDLNSCMPLGMSKAISDAQITASSYFT 1998
QY 2217 NMFATWSPKARLHLOGSNWAPQVNNKELQVDFQKTVKVTGVTTCQVKSLLTSMVY 2276
Db 1999 NMFATWSPKARLHLOGSNWAPQVNNKELQVDFQKTVKVTGVTTCQVKSLLTSMVY 2058
QY 2277 KEFLISSQDGHQWTLFTQNGKVKVFQGNQDSTFPVNVNSLDPLLRILRIHPQSVMHQI 2336
Db 2059 KEFLISSQDGRWTLFLQDGHTKVFGNQDSTFPVNVNSLDPLLRILRIHPQSVMHQI 2118
QY 2337 ALRMEVLGCEAODLY 2351
Db 2119 ALRMEVLGCEAODLY 2133
```

RESULT 15

US-10-131-510A-37

```
; Sequence 37, Application US/10131510A
; Publication No. US20030166536A1
; GENERAL INFORMATION:
; APPLICANT: Jellar, John S
; TITLE OF INVENTION: Modified Factor VIII
; FILE REFERENCE: 75-95J
; CURRENT APPLICATION NUMBER: US/10/131,510A
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: U.S. 09/315,179
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: U.S. 09/037,601
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: U.S. 08/670,707
; PRIOR FILING DATE: 1996-06-26
; PRIOR APPLICATION NUMBER: PCT/US97/11155
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: PCT/US94/13200
; PRIOR FILING DATE: 1994-11-15
; PRIOR APPLICATION NUMBER: U.S. 08/212,133
; PRIOR FILING DATE: 1994-03-11
; PRIOR APPLICATION NUMBER: U.S. 07/864,004
; PRIOR FILING DATE: 1992-04-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2133
; TYPE: PAT
; ORGANISM: Sus scrofa
; US-10-131-510A-37

Query Match 64.9%; Score 8057; DB 14; Length 2133;
Best Local Similarity 66.9%; Pred. No. 0;
Matches 1575; Conservative 209; Mismatches 345; Indels 226; Gaps 17;

QY 1 MQEISTCEFLCLLRPCFSATRRYYLGAVELSWDWMQSD-LGELPVDARPPRPVPSKSPFF 59
Db 1 MQEISTCEFLCLLRPCFSATRRYYLGAVELSWDWMQSD-LGELPVDARPPRPVPSKSPFF 60
QY 60 NTSVVYKKTFLVFETDHLFNIAKPPPMWGLLGPITQAEVYDVVITLKQMAHSPVSLHA 119
Db 61 GFSVLYKKTFLVFETDHLFNIAKPPPMWGLLGPITQAEVYDVVITLKQMAHSPVSLHA 120
QY 120 VGVSVWKASGEAYDDQTSQREKEDDKVFPQGSHTYVWQVLKENGFMASDPCLITYSL 179
Db 121 VGVSVWKASGEAYDDQTSQREKEDDKVFPQGSHTYVWQVLKENGFMASDPCLITYSL 180
QY 180 HYDLVKDLNSGLIGALLVCREGLSLAKETOTLEKILLFAVDEGKSWHSEYNSIMQDR 239
Db 181 HYDLVKDLNSGLIGALLVCREGLSLAKETOTLEKILLFAVDEGKSWHSEYNSIMQDR 240
QY 240 DAASARAMPKMTVANGYVNRSLPGLIGHRKEVYVHVIGMTTPEVHSIFLEGHFTFLVRN 299
Db 241 DPAPARAAQPMHTVANGYVNRSLPGLIGHRKEVYVHVIGMTTPEVHSIFLEGHFTFLVRN 300
QY 300 HQQASLETSPITFLTAQTLMLDGLQFLSCHTSSHQHGMWAYKVDSCPEEPQIMKNN 359
Db 301 HQQASLETSPITFLTAQTLMLDGLQFLSCHTSSHQHGMWAYKVDSCPEEPQIMKNN 360
QY 360 EEAEDYDDDLTQSEMDVVRFDQNSPFSQIRSVAKKPKTWVHYIAAEEEDWDYAPLV 419
Db 361 EEAEDYDDDLTQSEMDVVRFDQNSPFSQIRSVAKKPKTWVHYIAAEEEDWDYAPLV 419
QY 420 APDRSRYKSOYLNNGPQRTGRKYKVRFMAYYDETFKTRATQIHESGILGPLLYGEVGT 479
Db 420 SPDSRSKSLYLNNGPQRTGRKYKVRFMAYYDETFKTRATQIHESGILGPLLYGEVGT 479
QY 480 LLIIIFNKAQSRPNYIPHGITDVRPLYSRRLPKGVKHLKDFPILPCEIFKYKWTVDEG 539
Db 480 LLIIIFNKAQSRPNYIPHGITDVRPLYSRRLPKGVKHLKDFPILPCEIFKYKWTVDEG 539
QY 540 PTKSDPRCLTRYVSSPFAVNERDLASGLICPLICYKESVDQKGNQIMSKNVLPSVD 599
Db 540 PTKSDPRCLTRYVSSPFAVNERDLASGLICPLICYKESVDQKGNQIMSKNVLPSVD 599
```

[illegible]

Search completed: April 13, 2004, 14:20:04
Job time : 74 secs

This Page Blank (uspto)

This Page Blank (uspto)

This Page Blank (uspto)